RICE UNIVERSITY

Computational Discovery of Metal-Organic Frameworks with High Gas Deliverable Capacity

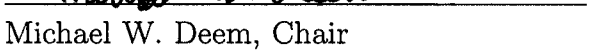
by

Yi Bao

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

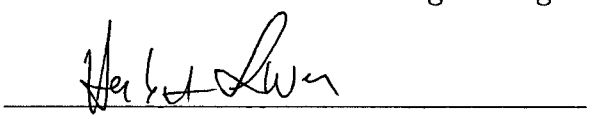
Doctor of Philosophy

Approved, Thesis Committee:



Michael W. Deem, Chair
John W. Cox Professor of Biochemical
and Genetic Engineering and Professor of
Physics & Astronomy

Douglas Natelson
Professor of Physics and Astronomy,
Electrical and Computer Engineering &
Materials Science and NanoEngineering



Herbert Levine
Hasselmann Professor of Bioengineering
and Director of Center for Theoretical
Biological Physics

Houston, Texas

April, 2017

ABSTRACT

Computational Discovery of Metal-Organic Frameworks with High Gas Deliverable Capacity

by

Yi Bao

Metal-organic frameworks (MOFs) are a rapidly thriving class of nanoporous materials with tunable chemistry and diverse applications in gas storage, gas purification, catalysis, sensing, and drug delivery. Efforts have been made to develop new MOFs with desirable properties both experimentally and computationally for decades. To guide experimental synthesis, we here develop a computational methodology to explore the chemical space of MOFs. I focus on MOFs with high gas deliverable capacity. This *de novo* design procedure applies known chemical reactions, considers synthesizability and geometric requirements of organic linkers, and efficiently evolves a population of MOFs to optimize a desirable property. We identify 48 MOFs with higher methane deliverable capacity at the Department of Energy standard condition of 65–5.8 bar loading–delivery than the well-known MOF-5 reference material in nine networks. In a more comprehensive work, we predict two sets of MOFs with high methane deliverable capacity at the 65–5.8 bar loading–delivery condition or the Department of Energy alternative standard condition of 35–5.8 bar loading–delivery. We also optimize a set of MOFs with high methane accessible internal surface area to investigate the relationship between deliverable capacities and internal surface area. This methodology can be extended to MOFs with multiple types of linkers and mul-

tiple types of SBUs.

Flexible MOFs may allow for sophisticated heat management strategies and also provide higher gas deliverable capacity than rigid frameworks. I investigate flexible MOFs, the MIL-53 family of materials, and Fe(bdp) and Co(bdp) analogs, to understand the structural phase transition of flexible frameworks. I derive a formula of the heat of gas adsorption in a flexible MOF. The challenges of simulating a system with a flexible host structure and incoming guest molecules are discussed. Preliminary results from isotherm simulation using the hybrid MC/MD simulation scheme on MIL-53(Cr) in the osmotic ensemble are presented. Suggestions for proceeding to understand the free energy profile of flexible MOFs are provided.

Acknowledgments

I would like to thank my advisor Dr. Michael W. Deem for his support and enlightenment for five years, without whom the thesis would not even be possible. I have been impressed by his pursuit of excellence, professional expertise in a wide spectrum of subjects, pragmatic behavioral style and charismatic personality. Besides, the most important point is that he has been showing high academic integrity in teaching and scientific research continuously. These traits will stand as my lasting references for the future.

Besides, I am grateful to Dr. Frits Daeyaert for his constant feedback and instruction that help solve my technical problems. And I want to express gratitude to my friends and colleagues, Dr. Pu Han, Dr. Man Chen, Dong Wang, Dr. Xi Li, Melia Bonomo, Fengdan Ye, and Shubham Tripathi for their intellectual support. I also appreciate Dr. Douglas Natelson and Dr. Herbert Levine as my committee.

I would like to thank my cousin, Yu Bao, for her genuine and sophisticated insights as a coeval. When I am writing this acknowledgment, she is in a tiny life slump when confusion and anxiety are weakening her capability of forecasting the future. But I believe self-reflection and mutual sympathy would help to overcome temporary difficulties as they always do.

Finally, I would like to thank my parents, Wenqing He and Xiaojun Bao, for their selfless and sustaining love. They provide me an endless source of sense of security as an intact family wherever I am physically. The liberty and respect they always devote to me strengthen my courage of challenging the unknown.

All above serves as my acknowledgments.

Contents

Illustrations

Tables

Chapter 1

Introduction

This thesis is composed of five parts. The first part is this chapter of general background of metal-organic frameworks and the scope of questions we are interested in. The second part of this thesis is an in-depth explanation of the metal-organic framework geometry and the methodology of the evolutionary algorithm in Chapters 2 and 3. The third part of this thesis is the discovery of rigid metal-organic frameworks by optimizing methane deliverable capacity or methane accessible surface area in Chapters 4 and 5. The fourth part of this thesis is the introduction of flexible metal-organic frameworks, heat of adsorption, the challenges that flexibility brings about, and the results of simulation in Chapters 6 and 7. The fifth part of this thesis is a conclusion and outlook in Chapter 8.

1.1 Metal-Organic Frameworks

Metal-organic frameworks (MOFs) are a group of rapidly growing materials first synthesized in 1995 [1]. They consist of metal ions or metal clusters, termed secondary building units (SBUs), and organic ligands, termed as linkers. MOFs have attracted researchers' interest because of their high surface area, large number of candidate constituents, and structural tunability. MOFs are of interest as an emerging class of nanoporous materials in the Materials Genome analogous to the Human Genome Project [2]. They are representative of crystalline structures in the Genome. MOFs

show exceptional potential in a variety of industrial applications [3], such as gas storage [4, 5], gas purification [6, 7, 8], catalysis [9, 10], drug delivery [11] and sensing [12, 13, 14]. MOFs are promising alternative candidates for clean energy storage [15]. Decades of efforts have been made to synthesize MOFs experimentally to search for particular structures that exhibit desirable properties. As a result, the total number of synthesized MOFs exceeds several thousand [16]. Computational simulations have further enlarged the number of potential MOFs to hundreds of thousands [17].

1.2 Experimental Synthesis and Computational Prediction

Experimental synthesis of MOFs is a relatively young and prospering field. Typical synthesis approaches [18, 19] include solvothermal synthesis [20], electrochemical synthesis [21], mechanochemical synthesis [22], microwave synthesis [23], and sonochemical synthesis [24, 25]. Synthesis protocols vary in the choices of solvents, solution conditions, and temperature to facilitate self-assembly of SBUs and linkers. While most MOFs are synthesized for laboratory scale, large-scale synthesis technologies are being developed for some MOFs [18], including HKUST-1 [26, 27], $[Mg(O_2CH)_2]$ [28], and MIL-53(Al) [28].

However, experiments may not synthesize and test all, or even a considerable portion of, possible structures due to the limit of overall synthesis capability and time. The chemical space of potentially interesting structures we are looking at is enormous. The aid of *in silico* simulations is important for predicting a finite number of favorable MOF structures that experimentalists may synthesize preferentially.

Current computational methods to predict new structures generally can be categorized into three types. One approach is to screen all plausible combinations of known candidate organic linkers and SBUs into a certain network (net) from given

libraries [29, 30, 17, 31]. This category of methods has presented an interesting landscape of new MOF structures and expanded the size of MOF databases. The second approach is to apply an evolutionary algorithm to explore the chemical space to optimize a desired property and avoid the computational cost of the brute-force method [32, 33, 34, 35]. This method sacrifices enumeration but accelerates the exploration in comparison with direct screening, which suffers from a combinatorial explosion. The third approach is a state-of-the-art application of machine learning algorithms to predict advanced materials [36, 37]. This category of methods is different from traditional simulations and data-driven approaches by learning from known experimental results and structural properties to predict high-performing candidates without explicitly confining the search criteria.

We have developed a *de novo* evolutionary algorithm to explore the composition space of linkers to optimize the gas deliverable capacity of predicted MOFs. As mentioned above, the current computational methods use compound libraries with known linkers. Our algorithm in general falls into the second category. The approach efficiently samples the linker library and performs random chemical transformations upon a pool of improving linker candidates. The algorithm keeps track of the synthetic route of each proposed linker with the information of chemical reactions it underwent and corresponding reactants. This methodology not only generates a population of MOFs with desired property but also provides synthesis guidance. It is worth noting that this method is inspired by the biased Monte Carlo method [38] and a similar application in drug design where *in silico* predictions suggested more than 55 lead compounds [39].

1.3 Gas Adsorption

As stated above, MOFs have been explored as potential materials for gas adsorption. The property we are interested in optimizing is the gas adsorption of MOFs. Particularly, we investigate MOFs as storage materials for compressed natural gas (CNG), mostly comprised of methane, in small vehicles. The Advanced Research Projects Agency-Energy (ARPA-E) of the US Department of Energy (DOE) set a target of 315 volumetric loading for methane at standard temperature and pressure (STP) with a loading pressure of 65 bar and a delivery pressure of 5.8 bar. The delivery pressure is higher than atmosphere pressure, i.e. 1 atm, because the gas needs the pressure to be pumped at pre-combustion pressure into vehicle engines. In this pressure range, which simulates the realistic condition of a vehicle engine, MOFs typically adsorb more gas at higher pressure than at lower pressure at ambient temperature. The adsorption difference between the loading pressure and delivery pressure is defined as the deliverable capacity, as detailed in the following chapter, which represents the usable amount of gas in a real vehicular application.

Initially, we implemented our evolutionary algorithm to predict MOFs with high methane deliverable capacity at 65–5.8 bar and 298 K in 9 distinct nets. We discovered 48 predicted MOFs in 4 nets that surpass a signature MOF, MOF-5, in terms of deliverable capacity at 65 bar. A further question of material chemistry is how structural characteristics correlate with property. Therefore, in a more comprehensive work, we optimized MOFs for three properties. 1) we evolved a population of linkers to optimize a measure of methane accessible surface area for each net; 2) we evolved a population of linkers to optimize deliverable capacity at 65 bar loading pressure, 5.8 bar delivery pressure, and 298 K for each net; and 3) we evolved a population of linkers to optimize deliverable capacity at 35 bar loading pressure, 5.8 bar

delivery pressure, and 298 K for each net. Many MOFs with high methane deliverable capacity are predicted. However, the highest deliverable capacity at 65–5.8 bar still does not pass the ceiling of 200 v(STP)/v previously encountered [40, 2, 41]. We investigated the correlation between deliverable capacity and accessible surface area and concluded that the correlation is network dependent.

1.4 Framework Flexibility

So far, our scope has been confined to rigid framework, and this is a prevalent assumption in the literature. That is to say, the framework acts as a static host structure without responding to any changes of external environment. MOFs do not always maintain a conformational rigidity under a mild reaction condition or even simply as-synthesized over time [42]. The "gate opening" or "breathing effect" phenomena are not uncommon in the MOF family, and examples include the MIL-53 variants [43, 44, 45].

A new group of MOFs, Fe(bdp) and Co(bdp) analogs, have been synthesized by Long's group which not only show an interesting phase transition between expanded phase at high pressure and collapsed phase at low pressure, but also exhibit a beneficial heat management during the transition [46, 47]. The structural phase transition reduces the heat of adsorption that is usually overlooked but is a significantly practical problem in new material development.

This structural phase transition potentially results in an increase in deliverable capacity compared to typical MOFs that have a type-I isotherm [48]. As mentioned, a recent publication reports that a spectrum of representative nanoporous materials, including MOFs, zeolites, zeolitic imidazolate framework (ZIFs), and porous polymer networks (PPNs), have a peculiar ceiling at 200 v(STP)/v of methane deliverable

capacity at 298K [2]. This value is well below the target set by ARPA-E and the question is whether the target may be reached. A limitation of nearly all computational evaluations to date is that they are performed on rigid frameworks. For instance, when adsorption of methane molecules is simulated, grand canonical Monte Carlo (GCMC) is typically performed where MOFs are treated rigid at the different gas pressures. It is known that such a simulation is a first order approximation to what really happens in the experiment.

The discoveries of a number of flexible MOFs inspire us to explore potential flexible MOFs that have an intrinsic phase transition, lower heat of adsorption, and greater deliverable capacity. The flexibility of MOFs poses challenges to computational simulation. At least two main questions arise. One is what force field should be applied for flexible MOFs. The other is what simulation scheme should be used to evaluate gas adsorption in the presence of a phase transition. These will be discussed in Chapter 6 and 7, where some preliminary results will be presented.

Chapter 2

Metal-Organic Frameworks

2.1 Overview

Since the archetypal metal-organic framework (MOF), MOF-5, was synthesized by linking zinc cationic clusters with benzene dicarboxylate molecules [49], interest in synthesis and design of MOFs has grown exceedingly rapidly in the past two decades. Even though MOF-5 was not the first MOF ever synthesized, it set off a wave of synthesizing organic-and-inorganic molecular complexes. Before its synthesis, the term "metal-organic framework" had not been introduced to academia. A few porous crystalline structures existed before the synthesis of MOF-5 that satisfy the general definition of what is now known as a MOF. It is especially encouraging that synthesizing a targeted composition of chosen organic compounds and inorganic clusters in a certain topology is now possible in the laboratory.

The definition of a MOF is commonly agreed as a crystalline coordination network that consist of two functional components. One is a metal or metal ion cluster that serves as a tether node. They are also termed as secondary building blocks (SBUs), because metal ions aggregate with oxygen and carbon atoms, as M–O–C clusters, to act as relatively robust molecular complexes [50]. In earlier research, "SBU" has been used to denote both organic units, the linker molecules, and inorganic units, the metal-containing blocks [51]. However, SBU is now predominantly being used for metal clusters only, and I am adopting this convention in this thesis. The other

component is the ligand molecule connecting the nodes, which are usually linear organic molecules with connecting sites, for example dicarboxylic groups. They are termed as organic linkers in the sense that they combine metal clusters together and form an open periodic framework in three dimensions. Topologically, SBUs and linkers can be abstracted as nodes and edges from a perspective of three dimensional graph.

MOFs feature large surface area, great pore size, and structural versatility. They appear promising for carbon-centered energy storage challenge [52]. Based on a recent statistical review of worldwide energy usage, over 80% of the energy consumed globally comes from fossil fuels, and the carbon dioxide released into atmosphere is greater than ever in history [53]. MOFs are a rising group of candidate crystalline structures for methane [41, 54] and carbon dioxide storage [55, 56, 57], and they may reshape the technology landscape of utilizing clean energy. The computational discovery of MOFs with high gas adsorption performance is one of the main topics in this thesis.

One may wonder how many unique structures of MOF have been synthesized and how many hypothetical MOF structures have been computed. This is a tricky question to answer because the definition of a MOF employed by researchers varies depending on the context of research. Databases of MOF structures and databases of general crystallography structures containing MOFs exist. MOFomics is an automated computational method for characterizing the porous networks of MOFs, and a database of MOF structures also bears this name [58]. The hypothetical metal-organic frameworks (hMOFs) database contains MOF structures generated by an enumerative screening method with a selected set of building blocks [17]. The computation-ready, experimental (CoRE) MOF database contains MOFs that are immediately available for molecular simulation [31]. And notably, the Cambridge Structural Database

(CSD) that is a leader in experimentally determined organic and metal-organic crystal materials also contains a large number of MOF structures [59]. The Crystallography Open Database (COD) is another collection of crystal structures including MOFs [60]. It is important to note that these databases have significant overlap, and they do not aim to be exclusive from each other. For example, MOFomics explicitly contains many MOFs from CSD; CoRE MOF database is strictly a subset of CSD that tailors selected MOF structures to enable high-throughput computation. Additionally, within the CSD database, a particular MOF structure may appear in several entries with different solvent molecules present. Therefore, one needs to apply caution when extracting data from these for computational purposes.

In 2011, there were only about 6,000 MOF structures in the CSD. In 2017, there are 61,036 MOFs based on a search with simple selection criteria in version 5.38 of the CSD. This constitutes about 7% of the CSD. The number of MOFs has increased much faster than the number of overall crystal structures in the CSD in the past decade [16].

2.2 Network

As discussed in Section 2.1, I used the concepts of nodes and edges to define MOFs due to the nature of framework configurational characteristics. SBUs and linkers stand for two classes of relatively rigid and independent components. These mirror the concepts of network topology or graph theory in computer science where various elements are arranged following a specific pattern. And consequently, terms “topology” and “network” have been introduced to describe the arrangement of SBUs and linkers in frameworks interchangeably. I am adopting “network” or simply “net” to make this thesis consistent, as this terminology is becoming predominantly used.

This analogy, however, leads to great confusion as physicists and chemists use terms from graph theory with different understanding. Olaf Delgado-Friedrichs, Michael O'Keeffe, Omar M. Yaghi and their colleagues formalize the terminology of crystal nets specifically in the context of MOFs [61, 62, 63]. Each net is represented by either a three-letter word or a three-letter word with an extension, and usually made boldface in formal written language. For example, **dia** stands for a commonly seen diamondoid crystal net. The assembly of SBUs and organic linkers following an abstract net is conceptualized as tiling where space is considered as polyhedra. This contribution facilitates the development of universal nomenclature and classification.

The networks indicate the complexity in the chemical space of MOFs. One net may allow multiple types of SBUs to be formed into one framework. For example, a **ntt** net allow two types of Cu-containing SBUs: one is the Cu_2 paddlewheel and the other is the Cu_3 SBU [62]. And one SBU may be a suitable candidate for multiple types of nets. For example, a Zr_6 metal node can be 6-connected to form a **spn** net, 8-connected to form a **csq** net, 10-connected to form a **bct** net or even 12-connected to form a **fcu** net [64].

As far as I know, there are two important databases of crystal nets. One of them is the Reticular Chemistry Structure Resource (RCSR) database that contains three dimensional nets, two dimensional nets, and polyhedra [62]. The nets of this database come from screening of known chemical compounds and enumerations. The other is the Euclidean Patterns in Non-Euclidean Tilings (EPINET) database that contains over 10,000 nets covering a broad spectrum of three dimensional Euclidean space [65]. Nets are generated by projecting tilings of a two dimensional hyperbolic plane onto three dimensional periodic surface in Euclidean space [66].

Chapter 3

Evolutionary Algorithm

3.1 Overview

The evolutionary algorithms are a class of optimization algorithms widely used in computer science and applied mathematics. They are inspired by natural selection in biology and aim to generate desirable solutions by optimizing a certain measure of fitness efficiently. In the context of computational chemistry, they are frequently referred as genetic algorithms, which are essentially the most popular subtype of evolutionary algorithms [67]. Even though the *de novo* evolutionary algorithm we develop keeps all the key features a typical one would have, such as initialization of population, mutation, and selection, it has a key merit that the synthetic routes of generated organic linkers are known. These synthetic routes are built upon a collection of chemical reactions and libraries of precursors that have been applied extensively in biochemistry and drug design. The precursors used in the synthetic routes are either chosen randomly from libraries of commercially available compounds or inherited from the current population.

The flowchart of the evolution algorithm is illustrated in Fig. 3.1. In summary, we use an evolutionary algorithm to evolve the MOF structures to optimize the methane deliverable capacity. We use a population of 100 linker molecules, which is initialized by compounds from a precursor library of commercially available molecules. Up to five chemical addition reactions may be applied to the linkers before they enter the

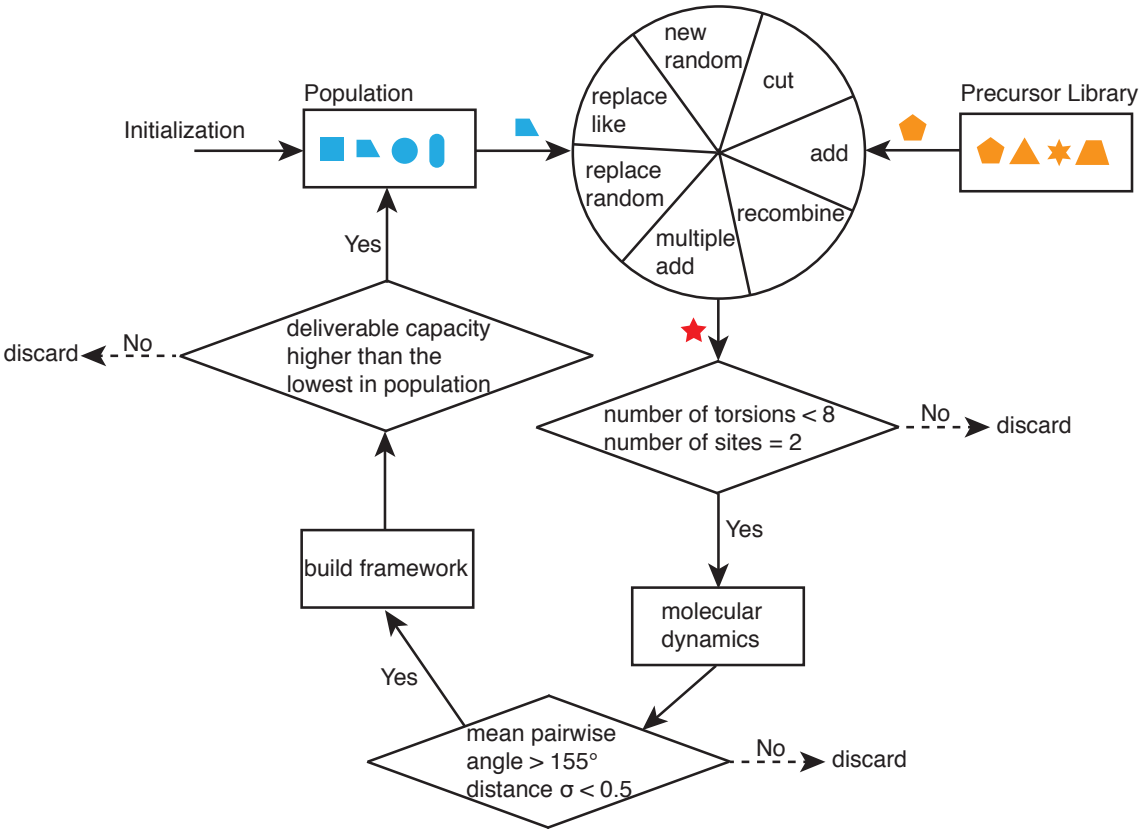


Figure 3.1 : Illustration of the evolutionary algorithm. The precursor library is used for initialization, and *add*, *multiple add*, *replace random*, *replace like*, and *new random* operations. One randomly chosen linker from the population undergoes one of seven evolution operations. The produced linker is evaluated by two filters first, i.e. number of torsions < 8 and number of sites $= 2$. If the linker passes these filters, it undergoes molecular dynamics to produce a set of conformations. These linker conformations are evaluated by the two additional filters of mean pairwise angle $\geq 155°$ and standard deviation of pairwise distance < 0.5 Å. If the linker passes all filters, it is used to build a MOF of the chosen network. Finally, if the constructed MOF has a greater deliverable capacity than the lowest one in the current population, the linker is inserted into the population in rank order.

initial population, because precursors directly drawn from the library may often lead to low deliverable capacity in the initial population. It is important to note that the population size remains 100 at all times after the initialization.

In each generation, one molecule undergoes one of seven evolution operations randomly, and produces a child molecule. One molecule is first randomly picked in the population. One of seven evolution operations is then uniformly and randomly chosen, and applied to the linker. If it is chemically impossible for the chosen operation to be applied to the linker, another operation will be chosen, until one is eventually performed. The evolutional operations are detailed in Section 3.3. It is possible using these operations to produce a molecule that has previously been created during the run. In this case, a new molecule from the population is picked, and the evolution step restarts. A hash log is used to record history of produced molecules and facilitate checking duplicates.

After a linker has gone through an evolution operation, selection determines whether it is reinserted to the population or discarded. The selection is based on a five-component vector with four binary filters, and one real number. These five components are evaluated sequentially as detailed in Section 3.4, as it is not necessary to evaluate further filters on a given proposed linker if one of the filters is not passed. The selection step operates hierarchically, comparing the first binary result, then the second, third, and fourth, and finally computing the deliverable capacity or accessible surface area.

After the final child molecule has satisfied all filters, it will be inserted into the population in the rank order if its deliverable capacity is higher than the current worst molecule. Since the population size remains constant, this results in the worst molecule being discarded. Otherwise, the child molecule itself is discarded. A run with

a given precursor library on a specific net performs 40,000 generations of evolutionary steps on proposed linkers and produces a final population of 100 evolved linkers.

For each proposed linker, before molecular dynamics is carried out, an initial configuration is determined. Linkers are first minimized in terms of their configurational energy in isolation, outside the MOF, to find a local energy minimum. This sets the bond lengths and bond angles to equilibrium values. Then amide bonds are set into the trans isomer when applicable, i.e. N-H being in a trans conformation with respect to C=O. This is done because molecules of the form NHRC(=O)R' will be planar in general, and the trans conformation is more stable. Therefore, the amide torsions are restricted. Further exploration identifies an approximately global energy minimum state by using the ant algorithm [68].

Molecular dynamics is carried out on each molecule to explore the flexibility and space of conformations to gauge their suitability as MOF linkers as detailed in Section 3.5.

A framework of the specified network is constructed for each conformation using the open source porous materials analysis suite Zeo++ [69] and model construction algorithm [70].

We use a chemical fingerprint, a type of Tanimoto similarity coefficient, to define similarity between two molecules. Each molecule can be uniquely encoded into a bit string in terms of chemical structure [71]. One string consists of 1092 bits, each denoting either absence (0) or presence (1) of a molecular motif. A motif is a selected single, double, triple or quadruple of types of atoms that are bonded, with types determined by the Merck Molecular Force Field. By calculating fingerprint similarities between each pair of two bit strings over all molecules, structural similarities can be quantitatively represented. Mathematically, fingerprint similarity t for two bit strings

r and s is defined as

$$t(r,s) = \frac{\sum_i r_i \wedge s_i}{\sum_i r_i \vee s_i}$$

where r_i and s_i are the ith bit of r and s, and $\wedge$ and $\vee$ are bitwise *and* and *or*.

3.2 Library

Two libraries of precursors have been used and compared in optimizing MOFs with either high gas deliverable capacity or surface area. The molecules to initialize the population of evolutionary algorithm are taken from these libraries. Besides, when *add*, *replace random*, *replace like*, *new random*, and *multiple add* operations are chosen, additional compounds are picked from them. Both libraries contain next-day available compounds from Sigma-Aldrich. Precursor library A, which contains additional compounds with exactly one carboxylic group and one benzyl group, has 42,284 molecules in total. These compounds can react to form dicarboxylic molecules that are particularly required by the nets we consider except **tbo**. Conversely, precursor library B, which contains additional compounds from Sigma-Aldrich that are used in medicinal chemistry or as organic building blocks, has 57,815 molecules in total. In Chapter 4 where I present the first results of optimizing MOFs with high methane deliverable capacity, precursor library B is used. In Chapter 5 where I present a more comprehensive work optimizing MOFs with either high methane deliverable capacity or accessible surface area, both libraries are used.

3.3 Evolutional Operations

1. *Add*. One chemically applicable reaction is randomly chosen from a list of 84 entries [72]. These reactions are adjusted from an original set of 70 reactions

[73]. If it is a zero-order reaction, no more reactants are needed; otherwise, if it is a first-order reaction, one more reactant is chosen randomly from the precursor library to join the reaction. Regardless of whether zero or first order, the reaction results in an addition to the chosen linker, and one reaction step is added to the end of synthetic route of the linker.

2. *Cut*. The last reaction step is removed from the synthetic route of the linker. In other words, the synthetic route of the linker is reversed by one step back toward its parent molecule.

3. *Replace random*. One reactant in the synthetic route is replaced by another from the precursor library that has suitable functionality for the reaction of that particular step.

4. *Replace like*. One reactant in the synthetic route is replaced by another that has the required functionality from the precursor library. In addition, the reactant and its replacement must be chemically similar within a tolerance. Chemical similarity is computed from a chemical fingerprint, discussed in detail below. The similarity threshold is set to 0.75.

5. *New random*. The linker is replaced by a new, randomly chosen compound from the precursor library. Up to five chemical addition reactions may be applied to the newly chosen linker before it enters the population.

6. *Recombine*. Another linker in the current population that shares at least one combinative node in the synthetic route with the linker to be evolved is randomly identified. The nodes denote where certain reactions with correspondent functionality groups occur. A child linker is created by choosing the initial part

of the synthetic route from the first linker and the final part of the synthetic route from the second linker. In the case of more than one common combinative node in the synthetic route, one node is randomly chosen.

7. *Multiple add.* One chemically applicable reaction is randomly chosen. When multiple reaction sites are possible on one molecule, the reaction takes place on all these sites within one operation. One reaction step is added to the end of the synthetic route of the linker.

3.4 Selection Criteria

1. *Number of torsions: binary filter.* The number of torsions is required to be no more than eight. Too many torsions cause excessive flexibility. The **tbo** net is an exception to this filter.

2. *Number of carboxylic sites: binary filter.* The number of carboxylic sites of the molecule is required to be exactly two, except for the **tbo** net which is given special treatment. The metal SBUs examined in this work may give rise to alternative networks when combined with organic linkers with more than two connection sites; here we restrict our search to the aforementioned networks and hence to linear two-connected linkers.

3. *Mean pairwise angle: binary filter.* After the first two filters are passed, a 30 ps molecular dynamics simulation at 298 K on the molecule is performed to produce a set of 300 conformations. The mean pairwise angle over all conformations produced from molecule dynamics is required to be larger than a cutoff. The angle is defined as that between two vectors, one orienting from the carbon atom and passing through the midpoint of two oxygen atoms for the first carboxylic

group metal attachment site and the other for the second carboxylic attachment site. It is anticipated that successful linkers will be linear, with anti-parallel attachment sites, and so desired angles are set as $\geq 155°$. The **tbo** net is an exception to this filter.

4. *Standard deviation of pairwise distance: binary filter.* To avoid an entropic penalty against incorporation of the linkers into the MOF structure, rigid linkers are preferred. Such rigid linkers are more likely to successfully self-assemble the MOF structure. To quantify rigidity, we measure the standard deviation of pairwise distance between linker sites. By calculating several typical linear organic linkers of SURMOF-2 [74], we find that the largest standard deviation of pairwise distance is 0.24 Å. In this study we used a conservative standard deviation of pairwise distance of 0.5 Å as our cutoff. The distance is defined according to the coordinates of the carbon atoms in each of the carboxylic sites. The standard deviation is evaluated over all conformations produced in the molecular dynamics run. The **tbo** net is an exception to this filter.

5. *Methane deliverable capacity: real.* If the linker passes the filters above, a MOF is built using the linker [69, 70]. The fifth component of the vector is a measure of methane deliverable capacity. We use grand canonical Monte Carlo to simulate equilibrium methane uptake process in given frameworks at temperature of 298 K, and to calculate the deliverable capacity as detailed in Section 3.6.

3.5 Molecular Dynamics

The molecular modeling software TINKER 6.3 [75] is used to perform the molecular dynamics, if this molecule satisfies the first two filters. The dynamics is computed for

30 ps at a temperature of 298 K. The Merck Molecular Force Field commonly used in calculations on compounds in organic chemistry is used here. Conformations are sampled every 0.1 ps, giving a total of 300 conformations for each molecule.

3.6 Grand Canonical Monte Carlo (GCMC)

Grand canonical Monte Carlo is implemented with RASPA [76] when I evaluate the deliverable capacity. The frameworks are treated as rigid host, temperature is fixed, volume of system is fixed, and the number of gas molecules in the system is allowed to change. Therefore GCMC is the appropriate ensemble to simulate the system. In each simulation, I construct a minimum supercell of the framework in three dimensions with an interaction radius cutoff of 12.5 Å. We use the universal force field (UFF) to calculate non-bonded interactions between framework and gas molecules, including a tail correction [77]. We calculate adsorption at the loading pressure of 65 bar and at the depletion pressure of 5.8 bar, both at 298 K. The second result is subtracted from the first to give the deliverable capacity, i.e. the net excess loading between the two pressures. The reported values are multiplied by RT_0/P_0 where $T_0 = 273.15$ K and $P_0 = 1$ bar to give units of v(STP)/v. To maximize efficiency, I perform one methane uptake simulation with 300 Monte Carlo steps in total for every 50 linker conformations. Greater numbers of Monte Carlo steps did not significantly affect the average predicted deliverable capacity. Linker molecules sampled over the molecular dynamics are used to build corresponding MOFs with given SBUs and topology. The final deliverable capacity of each MOF is averaged over the conformations.

Chapter 4

Optimize MOFs with High Methane Deliverable Capacity

4.1 Overview

In this work, we predict MOFs with high methane deliverable capacity in nine different networks, each formed with a specified SBU. For each network and chosen SBU, we evolve a population of linker molecules and build MOFs to optimize methane deliverable capacity. As will be described with details in the following, the set of linker molecules is initialized from a set of commercially available compounds. The linker molecules are evolved *in silico* by a set of chemical reactions, so that only chemically-synthesizable linkers are considered as the deliverable capacity is optimized. These *in silico* reactions are known to succeed with high probability and high yield [72].

Among the materials we predict are ones with a higher deliverable capacity than MOF-5 (i.e., IRMOF-1), a well known MOF at the time of this study with a high methane deliverable capacity. The experimentally measured capacity of MOF-5 is 177 v(STP)/v at 65 bar adsorption pressure, 5.8 bar delivery pressure, and 298 K [49, 78, 79]. We select nine well known networks, **acs**, **cds**, **dia**, **hxg**, **lvt**, **nbo**, **pcu**, **rhr**, and **sod**, to search for MOFs with a high methane deliverable capacity. We find 48 MOFs with a greater deliverable capacity than MOF-5: 14 in **acs**, 24 in **cds**, 2 in **nbo**, and 8 in **pcu**, with the best deliverable capacity found in **pcu** and 8% greater than that of MOF-5.

4.2 Deliverable Capacity

Deliverable capacity (DC) is a metric of how much gas a certain structure of nanoporous material may actually deliver under given working condition. In the context of this thesis, it is defined as the difference, reported in a volumetric unit, between the adsorption at an adsorption pressure and one at a delivery pressure. The adsorption pressure is either 65 bar or 35 bar while the delivery pressure is 5.8 bar. The loading pressure of 65 bar is used in this chapter. The reported deliverable capacity is multiplied by RT_0/P_0 where $T_0 = 273.15$ K and $P_0 = 1$ bar to give a unit of v(STP)/v, where STP stands for standard temperature and pressure.

Typically, gas adsorption climbs up an isotherm as the applied gas pressure increases. Under the assumption of rigid framework, these isotherms present type-I curves which suggest a monolayer Langmuir model would be appropriate to apply [48]. I will come back to this point in Section 4.3. Indeed, a high deliverable capacity suggests a particular MOF may serve as a great gas storage material in small vehicle for instance.

4.3 Results of Optimizing Deliverable Capacity at 65–5.8 bar Condition

The top deliverable capacity of each net is shown in Table 4.1, among which the **pcu** net gives highest deliverable capacity of 191.1 v(STP)/v. This value is 8% greater than our calculated value for MOF-5. As far as we know, predicted deliverable capacities significantly greater than MOF-5 under DOE target conditions have not been achieved before using any computational design approaches.

The distribution of deliverable capacities of all molecules in nine nets is shown

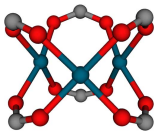
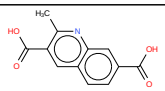
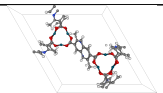
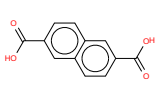
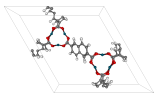
Net	Highest Deliverable Capacity v(STP)/v	Number of Candidates
acs	183.89	817
cds	189.64	512
dia	173.28	75
hxg	162.02	30
lvt	175.16	720
nbo	179.78	516
pcu	191.07	61
rhr	155.67	239
sod	171.73	16

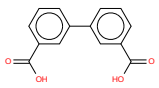
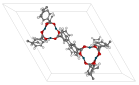
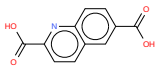
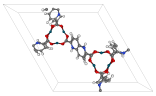
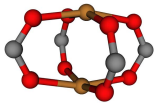
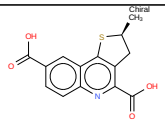
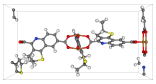
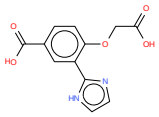
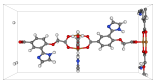
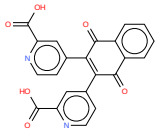
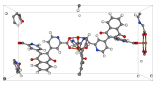
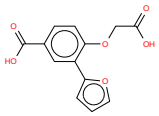
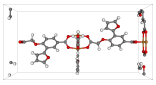
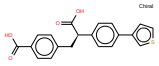
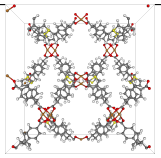
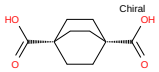
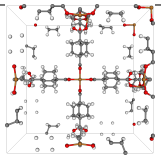
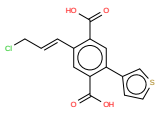
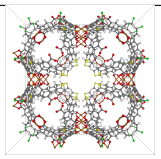
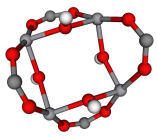
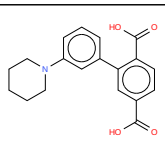
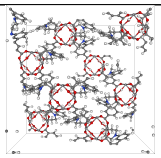
Table 4.1 : Top deliverable capacity for each net. Candidates are defined as all molecules that have passed all filters and have a positive deliverable capacity during the whole evolution. With our force field, MOF-5 has a deliverable capacity of 177.6 v(STP)/v. The best predicted MOF has a deliverable capacity of 191.1 v(STP)/v.

in Fig. 4.1. We find 48 MOFs, in four of the nine nets, having a higher deliverable capacity than MOF-5. We also show the structural formulas of linker molecules with highest methane deliverable capacity for respective nets in this figure.

Our calculation of the deliverable capacity of MOF-5 at these DOE target conditions is 177.6 v(STP)/v, nearly identical to the experimentally measured value of 177 v(STP)/v [79]. In **pcu** net alone we find eight linker molecules that outperform the 177.6 v(STP)/v methane deliverable capacity of MOF-5, as shown in Fig. 4.2.

Table 4.2 gives a few MOFs with the highest deliverable capacities in each of the nine nets. Both SBUs and linker molecules are shown.

SBU + linker = MOF	net	Deliverable Capacity (v(STP)/v)
  	**acs**	183.89
 		181.75

				181.21
				180.56
			**cds**	189.64
				189.47
				189.12
				185.59
			**lvt**	175.16
			**nbo**	179.78
			**rhr**	155.67
			**dia**	173.28

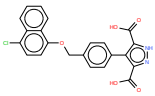
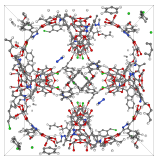
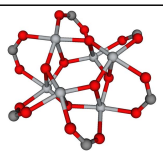
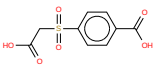
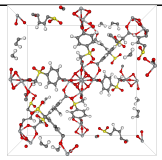
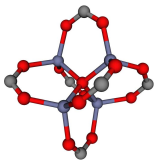
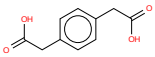
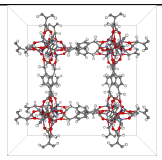
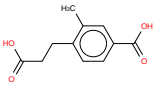
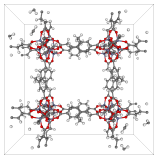
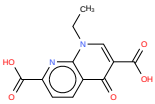
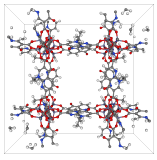
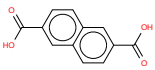
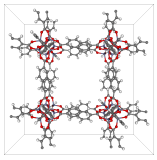
			**sod**	171.73
			**hxg**	162.02
			**pcu**	191.07
				183.74
				180.53
				179.98

Table 4.2 : A table of MOFs from predicted linker molecules and specified SBUs in the nine nets. Corresponding networks and calculated deliverable capacities are also listed. The SBUs are Pd_3 for **acs**; Cu_2 for **cds**, **lvt**, **nbo** and **rhr**; $V_4(OH)_4$ for **dia** and **sod**; Ti_6O_6 for **hxg**; and Zn_4O for **pcu**. Each SBU figure with O-C-O connecting sites is depicted in a color scheme. Red: oxygen atoms; dark gray: carbon atoms; white: hydrogen atoms; other colors: respective metal atoms.

We record the synthetic route of each linker molecule in the population, along with what reactants are introduced from the precursor library and what chemical

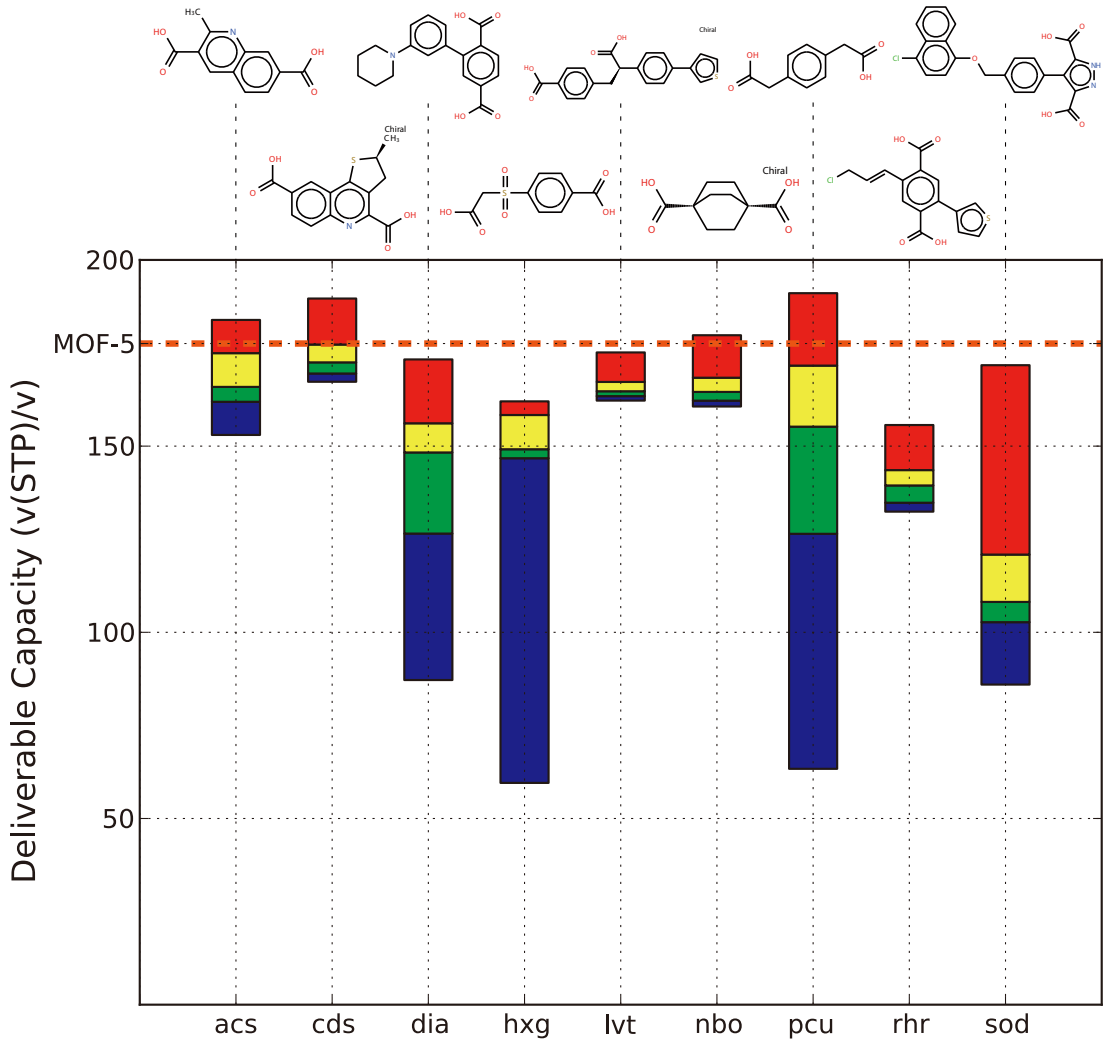
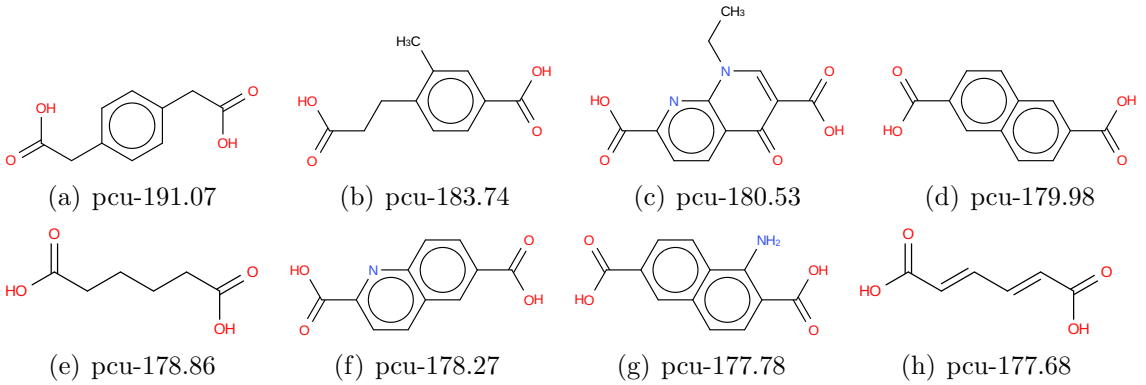


Figure 4.1 : Bar plot of deliverable capacity of all molecules in final populations over nine nets. Our method finds 48 linkers with better deliverable capacity than MOF-5, in the nets of **acs**, **cds**, **nbo**, and **pcu**. The difficulty of constructing valid frameworks varies for different networks, and 100 linkers satisfying all constraints are not found for **dia**, **hxg**, **pcu**, and **sod**, as shown in Table. 4.1. The structural formulas of best linker molecules for each net is illustrated on the top of each bar. The different colors each represent one quartile of the data.



(a) pcu-191.07 (b) pcu-183.74 (c) pcu-180.53 (d) pcu-179.98

(e) pcu-178.86 (f) pcu-178.27 (g) pcu-177.78 (h) pcu-177.68

Figure 4.2 : Representative structural formulas of the eight linker molecules in **pcu** net with higher deliverable capacity than MOF-5. Molecules are named as "net - deliverable capacity".

reactions are applied. One example of a synthetic route for a linker in the **cds** net is shown in Fig. 4.3.

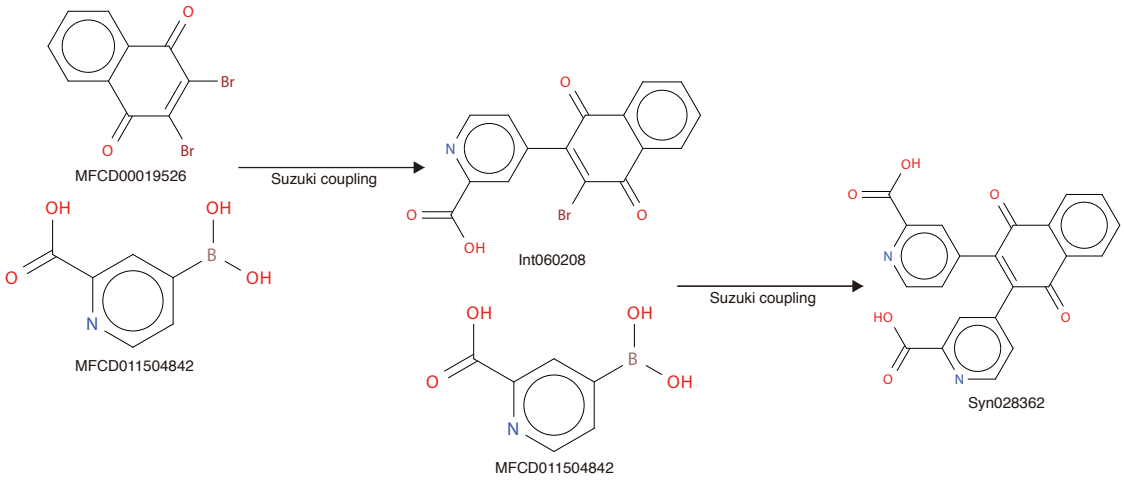


Figure 4.3 : Synthetic route of a proposed linker Syn028362 in **cds** net. MFCD00019526 and MFCD011504842 are picked from the precursor library. They undergo a Suzuki coupling to produce the intermediate molecule Int060208. Int060208 undergoes another Suzuki coupling with precursor MFCD011504842 to produce the final product Syn028362.

The deliverable capacity of 48 predicted MOFs in four nets outperformed that of

MOF-5. The one with highest deliverable capacity of each net is plotted in Fig. 4.4a. To characterize the diversity of the optimal MOFs, we apply a modified Langmuir adsorption model as given by Eq. 4.1 [80].

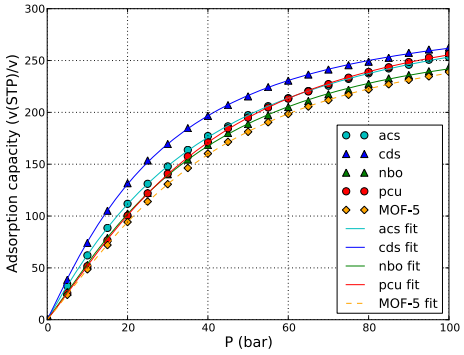
$$A = A_\infty \frac{P}{P + P_{1/2}} \left[1 - \theta \frac{P P_{1/2}}{(P + P_{1/2})^2} \right] \tag{4.1}$$

where P is pressure, A_∞ is the saturation adsorption capacity of a MOF when the pressure goes to infinity, $\theta < 0$ is a measure of methane-methane attractive interactions, and $P_{1/2}$ is the pressure at which adsorption capacity reaches one half of the maximum without the correction term, i.e. when $\theta = 0$. To calculate the Langmuir adsorption isotherm parameter values, adsorption at 21 evenly spaced points between 0 and 100 bar was calculated. The A_∞ and $P_{1/2}$ parameters were determined by fitting the Langmuir adsorption isotherm to these 21 data points by non-linear least-squares Marquardt-Levenberg algorithm.

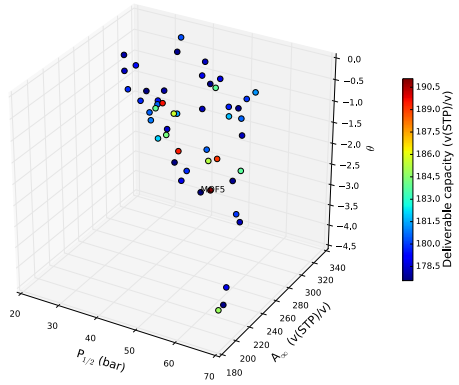
The fitted parameters of the modified Langmuir equation for optimal MOFs are illustrated in Fig. 4.4b. The Henry's constants, $H = A_\infty / P_{1/2}$, which characterize the low-pressure adsorption, are shown in Fig. 4.4c.

4.4 Discussion and Conclusion

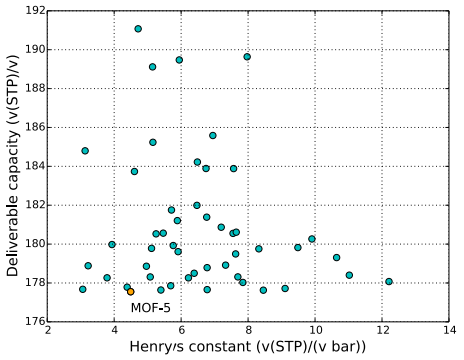
A guiding principle of our methodology is to efficiently search the space of possible MOF compounds. We explore the composition and conformation space of possible linkers subject to the known constraints of chemical synthesis and optimize structures with predicted high deliverable capacity. We find that a sampling rather than enumeration approach proves more successful, echoing the results in zeolite design [81].



(a) Methane adsorption isotherms



(b) Fitted parameters using the modified Langmuir isotherm model



(c) Henry's constant

Figure 4.4 : (a) Calculated methane adsorption isotherms of four predicted MOFs at 298 K. MOF-5 is also shown as a reference (dashed). (b) The MOFs concentrate in a narrow band in the parameter space. (c) MOFs with the highest deliverable capacity ($>$ 182 v(STP)/v) are found predominantly in the range of 3 to 8 v(STP)/v of Henry's constant.

MOF-5 is found among the results, identified in the **pcu** net as pcu-177.55. At the time this study was completed, MOF-5 and HKUST-1 were materials with long-standing records of the highest deliverable capacity, 177 and 182 v(STP)/v, respectively [79]. We also noticed that UTSA-76a was an additional MOF with high deliverable capacity, but along with HKUST-1, was not accessible in any of the nine networks we consider [82]. MOF-5 has been intensively studied and synthesized industrially via several approaches [49, 83]. We identified 48 MOFs with a predicted methane deliverable capacity greater than the well-studied MOF-5.

Studies in MOF design have identified structures with a variety of linker sizes, shapes, and chemistry. For instance, large porphyrins have been used as building blocks as MOF linkers [84]. Long aromatic hydrocarbons, with up to 11 connected phenylene rings, have been used to construct IRMOF-74-XI [16]. Numerous linkers for MTV-MOFs and MUF-1–MUF-7 also contain long aromatic hydrocarbons [85, 86]. This diversity of linker chemistry suggests that *de novo* designed linkers using the approach introduced here by *in silico* exploration of chemical space may also be successfully incorporated into realistic MOFs. In this regard, the recent success of *in silico* design of an organic structure directing agent for the zeolite STW is also encouraging [87].

Additional factors to be considered in experimental realizations of these predictions include solubility of a given linker in a relevant reaction solvent, variety of solvents, and temperature. Structural properties of the linkers influence the likelihood of MOF self-assembly. Rigid linkers are often preferred to avoid an entropic penalty for incorporation of linkers into the MOF. Consequently, we require the standard deviation of pairwise distance, calculated over the set of molecular conformations, to be less than 0.5 Å. We also apply a filter on the number of torsions to eliminate

linkers with excessive flexibility. As successful linkers are anticipated to be linear with anti-parallel attachment sites, we also apply the filter of mean pairwise angle. Loose filters tend to lead to more valid but mediocre candidates, while strict filters lead to fewer valid but outstanding candidates.

We notice that some evolved linker molecules have long side groups pendant to the backbone between the two site-connecting carboxylic groups. Intuitively, these side groups could help increase capacity because they provide more possible adsorption sites for gas molecules. The entropy of these side groups might inhibit MOF assembly if the configurational constraints on the side groups are significantly greater in the MOF than in solution. As long as these side groups are relatively unconstrained by SBUs, i.e. no collisions prevent framework construction, such side groups should not be notable barriers against MOF assembly. As the side groups become larger, they may prevent an efficient packing of methane. This situation is avoided by the optimization in the evolutionary linker search procedure. Additionally whether side groups actually increase deliverable capacity, and not simply adsorption at both 65 and 5.8 bar, is a subtle point that can only be answered by more detailed Monte Carlo or molecular dynamics calculations.

The parameters of the modified Langmuir adsorption model were used to characterize the performance of predicted MOFs with adsorption capabilities exceeding MOF-5. Interestingly, as shown in Fig. 4.4b, we predict that MOFs optimal for deliverable capacity are in a limited region of the parameter space. Since for a given $P_{1/2}$ and θ, larger deliverable capacities are produced by materials with larger A_∞, the optimal materials shown in Fig. 4.4b are those with the largest A_∞ values at each $P_{1/2}$ and θ.

In short, we predicted MOFs that optimize methane gas deliverable capacity in

nine networks, using a new *in silico* linker evolution procedure. Notably, 48 predicted linkers in the **acs**, **cds**, **nbo**, and **pcu** networks surpass MOF-5, a signature MOF with good methane deliverable capacity in a 65–5.8 bar pressure swing system at 298 K. For each of the nine networks, we used an SBU that was known experimentally to provide the appropriate linker attachment sites. The predicted linkers were constrained to be rigid and to correctly form a MOF geometry. The predicted linkers have promising synthetic feasibility, as previous experience in drug design with the *in silico* reactions considered here succeed in high yield 86% of the time [73]. Clearly the predicted linkers are just one part of the puzzle, as practical sythesis conditions must still be identified for the linker and SBU to self-organize into the MOF.

Chapter 5

Optimize MOFs with High Methane Deliverable Capacity or High Internal Surface Area

5.1 Overview

This study extends the previous one presented in Chapter 4 by using one more net **tbo**, an additional precursor library, and optimizing two more quantities of deliverable capacity at 35–5.8 bar and methane accessible internal surface area. The new net **tbo**, one example of which is HKUST-1, incorporates three-connected linkers rather than two-connected ones. The new precursor library, denoted as library A, contains additional compounds than the previously used one, denoted as library B. Two more quantities are optimized to explore the correlation between deliverable capacities and surface area.

In this study, we use the evolutionary algorithm to automate the identification of linker designs which lead to MOFs with optimized certain properties, i.e. methane accessible surface area or deliverable capacity [32]. We have assessed synthetic feasibility of organic linkers, and the synthetic routes of them typically include fewer than five chemical reaction steps. This procedure generates a population of predicted organic linkers by *in silico* evolution. We consider two different precursor libraries containing $\sim 50,000$ compounds each and compare their performance. We apply 84 chemical reactions that have been successfully applied in drug design [73]. This approach has been used to predict chemically synthesizable organic structure directing

agents for zeolites [72], and the method has successfully predicted an organic structure directing agent for synthesis of the zeolite STW, a very challenging synthesis problem [87].

As before, we consider the nine well-known MOF nets **acs**, **cds**, **dia**, **hxg**, **lvt**, **nbo**, **pcu**, **rhr**, and **sod**, which constitute some of the possible regular, semiregular, and minimal nets [63] based on combining selected SBUs with linear, two-connected linkers. We also consider the **tbo** net constructed from three-connected linkers. Three sets of results are obtained using the evolutionary algorithm: 1) we evolve a population of linkers to optimize a measure of methane accessible surface area for each net; 2) we evolve a population of linkers to optimize deliverable capacity at 65 bar loading pressure, 5.8 bar delivery pressure, and 298 K for each net; and 3) we evolve a population of linkers to optimize deliverable capacity at 35 bar loading pressure, 5.8 bar delivery pressure, and 298 K for each net. The **tbo** net is considered only in set 3. For set 2, results for the nine symmetric networks at 65–5.8 bar have been previously presented [32]. The results presented here for set 2 are novel and complement to those previously presented. We point out that in comparison to the previous work [32], the three sets of results presented here are generated using the same algorithm but at a larger scale, i.e. by optimizing deliverable capacities at two pressure conditions and surface area separately. Also, one additional precursor library, library A, is applied to be compared with library B probing the differences of the performance.

Additionally, one more network, **tbo**, is considered in the optimization of deliverable capacity. For each set of results, we investigate the relationship between the measure of surface area and deliverable capacity. For example, even though set 2 is optimized for deliverable capacity at 65–5.8 bar, we also computed the deliverable capacity at 35–5.8 bar and the accessible surface area for each of the compounds

identified.

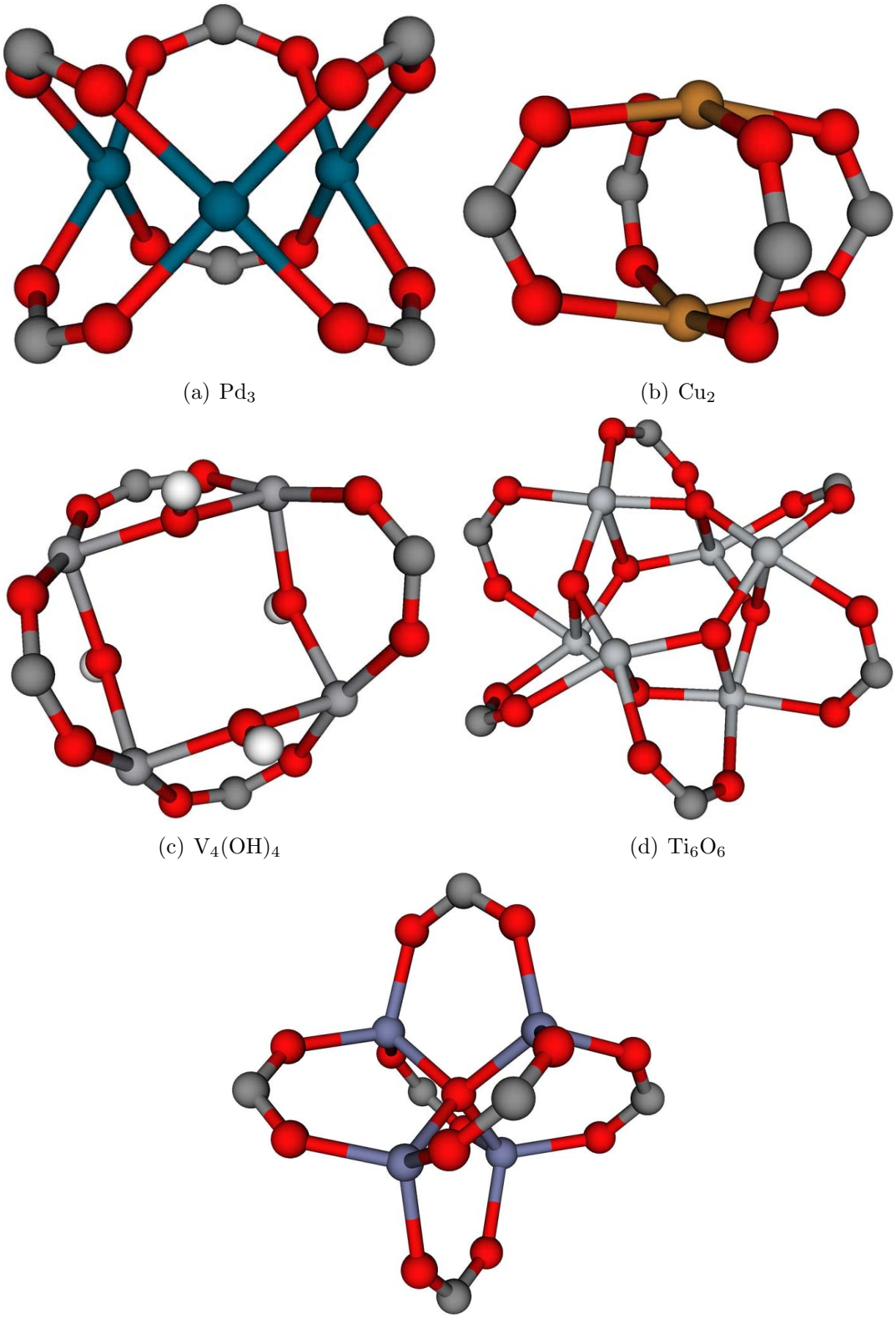
It is worth noting that the pairwise angle filter is applied to these conformations to require that the two carboxylic groups point in nearly opposite directions: for surface area optimization we require angles $\geq 155^\circ$ and for deliverable capacity optimization we require angles $\geq 160^\circ$.

To calculate surface area, a MOF is constructed from the proposed linker and the metal SBU in the chosen network with the software suite Zeo++ [69]. MOFs are crystalline materials, and thus we construct the unit cell of the MOFs from the metal SBU and organic linker as input based on the net information [70]. The metal SBUs are selected according to simplicity and charge balancing [88]. The SBUs chosen for each of the ten networks are shown in Table 5.1 and Fig. 5.1. The MOF is constructed by positioning the SBUs to distinct vertices of the net. After aligning the set of SBUs, the linkers are added to the framework, and the unit cell dimensions are scaled to accommodate the linkers.

Net	SBU	Description
acs	Pd_3	trigonal prism
cds, **lvt**, **nbo**, **rhr**, **tbo**	Cu_2	square "paddlewheel"
dia, **sod**	$V_4(OH)_4$	tetrahedral
hxg	Ti_6O_6	hexagonal
pcu	Zn_4O	octahedral

Table 5.1 : SBUs for the ten well-known MOF networks. One unique metal SBU is chosen for each network. SBUs are selected to be charge neutral.

When optimizing deliverable capacity, rather than surface area, we only use precursor library B. The minimum supercell to accommodate an interaction radius cutoff 12.5 Å is built from the SBU, linker, and network using the software suite Zeo++ [69]. Then we simulate methane gas adsorption with grand canonical Monte Carlo



(a) Pd_3 (b) Cu_2

(c) $V_4(OH)_4$ (d) Ti_6O_6

(e) Zn_4O

Figure 5.1 : Spatial configurations of secondary building units (SBUs) with O-C-O connecting sites. Red: oxygen atoms; dark grey: carbon atoms; white: hydrogen atoms; other colors: respective metal atoms.

for 300 steps. We calculate deliverable capacities every 50 MOF conformations over a set of 300 conformations at 298 K and two pressures, either 65 bar and 5.8 bar, or 35 bar and 5.8 bar.

In addition, since the **tbo** net requires each linker to have three connecting sites to the SBU, the filters to gauge the linker geometry are different than previously reported [32]. After the number of connecting sites is evaluated, the algorithm calculates the distances d_i, where $i = 1, 2, 3$, between each pairwise carboxylic groups over a set of 300 linker conformations. Before calculating the deliverable capacity, we require $\sigma^m/\overline{d_i^m} < 2\%$ for each conformation m, where σ^m is the standard deviation of the three distances of the linker for conformation m and $\overline{d_i^m}$ is the distance averaged over three pairwise carboxylic sites for conformation m. The purpose of this requirement is to demand rigidity of linker molecules. This filter also accommodates the symmetry assumption of framework building.

5.2 Methane Accessible Internal Surface Area

In a gas adsorption process, gas molecules interact with the host atoms to transform from gas phase into adsorbed phase. It is commonly argued that a porous material which exhibits large surface area will be able to provide more interaction sites, or binding sites, for guest molecules. It is also generally believed that high surface area should be positively correlated with good performance of materials in gas-related applications [89, 30, 90]. One widely used theory is Brunauer–Emmett–Teller (BET) theory that treats adsorption as a multilayer process and consequently a BET surface area can be properly defined [91].

In this work, instead, surface area is measured by the product of $GSA{\cdot}VSA$, where GSA (surface area per unit mass, m^2/g) denotes gravimetric surface area, and VSA

(surface area per unit volume, m^2/cm^3) denotes volumetric surface area. Generally, high GSA linkers tend to be long or branched, so as to maximize surface area with as much light-weight organic material as possible. Conversely, high VSA is usually the result of utilizing short linkers, so as to minimize the volume of material and increase the surface area by growing linker width. The product of GSA and VSA represents a compromise between these two competing effects [92] and has been used in study of hydrogen storage potential [93]. Intuitively, we may anticipate that the highest $GSA \cdot VSA$ will be the result of short organic linkers, to minimize material volume, with significant branching, to maximize the contribution of the organic to the surface area.

5.3 Results of Optimizing Methane Accessible Surface Area

MOFs were evolved to optimize the methane accessible surface area. This set of results contains 1793 MOFs. In this set of results, linkers for each of the ten MOF nets were designed using two different precursor libraries. The best 100 MOFs are retained for each net and precursor library. We compare the computed surface areas with a set of previous results which applies gradient-based optimization in an abstract shape-space of MOF linkers [92, 94]. Results are shown in Table 5.2. In most cases, the present evolution approach identifies linkers which produce greater $GSA \cdot VSA$ compared with values from previously identified optimal linker shapes; furthermore, the linkers in this work are chemically synthesizable following the suggested routes. Particularly favorable results are found for the nets of **acs**, **dia**, **nbo**, and **lvt**. The superior results achieved in this work indicate exploration of linker shapes that fall outside the bounds in the previous optimization work [92]. Previous work imposed symmetry constraints on the linkers, while in this work there is no constraint, and the

best results are typically achieved with asymmetrical molecules. The **sod** net with precursor library B gives the best instance of this, e.g. as shown in Fig. 5.6(c) and Fig. 5.7(i), where the very large $GSA \cdot VSA$ is achieved using a short linker with a long side group.

Net	Precursor Library A			Precursor Library B			$GSA \cdot VSA$ from
	GSA	VSA	$GSA \cdot VSA$	GSA	VSA	$GSA \cdot VSA$	previous results [92]
acs	4,468	3,094	13,826,100	4,980	2,838	14,132,400	10,440,000
cds	6,307	2,636	16,624,500	5,039	3,031	15,274,800	17,630,000
dia	5,625	2,498	14,053,700	5,894	3,202	18,870,200	10,065,000
hxg	4,740	1,971	9,342,410	4,945	1,848	9,137,620	10,560,000
lvt	5,929	2,389	14,165,000	6,104	2,475	15,107,800	5,795,000
nbo	5,641	2,366	13,347,100	5,285	2,416	12,769,400	10,875,000
pcu	5,558	2,244	12,471,900	3,897	2,687	10,469,300	11,500,000
rhr	5,823	1,501	8,737,070	4,675	1,953	9,131,470	8,030,000
sod	6,329	1,236	7,820,700	6,231	2,870	17,885,400	7,930,000
tbo				4,099	2,381	8,794,350	

Table 5.2 : Evolved best values of GSA, VSA, and $GSA \cdot VSA$ from two precursor libraries of commercially available compounds. Linkers were selected to optimize $GSA \cdot VSA$. Only results from linkers with pairwise angle $\geq 155°$ are shown. Previous results were calculated via gradient-based optimization in an abstract chemical shape space [92]. Precursor libraries A and B provide consistent results for **acs**, **cds**, **hxg**, **lvt**, **nbo**, and **rhr**. Also shown is the **tbo** net results from deliverable capacity at the 35–5.8 bar condition set using library B. GSA: m^2/g; VSA: $m^2/(cm^3)$; $GSA \cdot VSA$: $m^4/(g\ cm^3)$.

In Fig. 5.2(a), the GSA and VSA values for the final population of linkers in both precursor libraries and for each network are shown. We illustrate the MOFs with Pareto frontier in GSA–VSA space to identify the favorable nets. The Pareto frontier for each network is shown in Figs. 5.2(b)–5.2(d).

Some chemically promising linkers identified from both libraries are shown in Figs. 5.3–5.6.

Fig. 5.7 shows ten sample MOFs with predicted linkers for each net. Fig. 5.8 shows an example of synthetic route of molecule nbo-A-13347100.

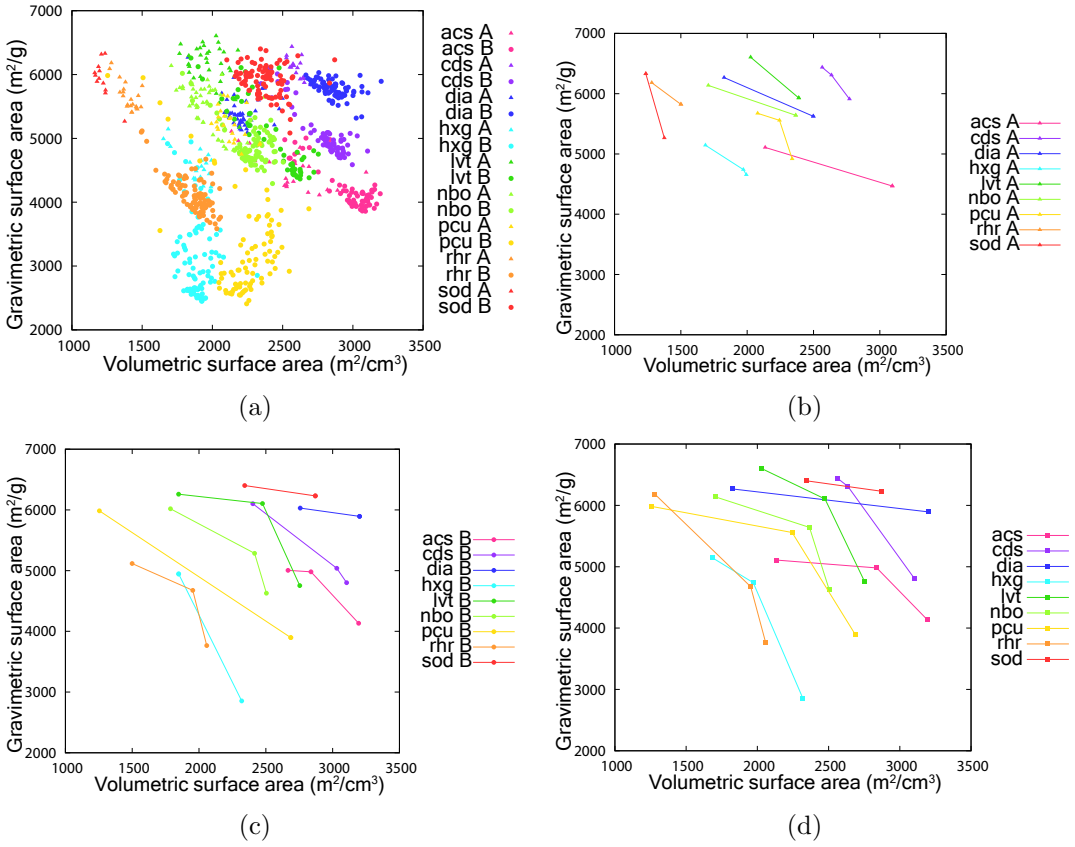
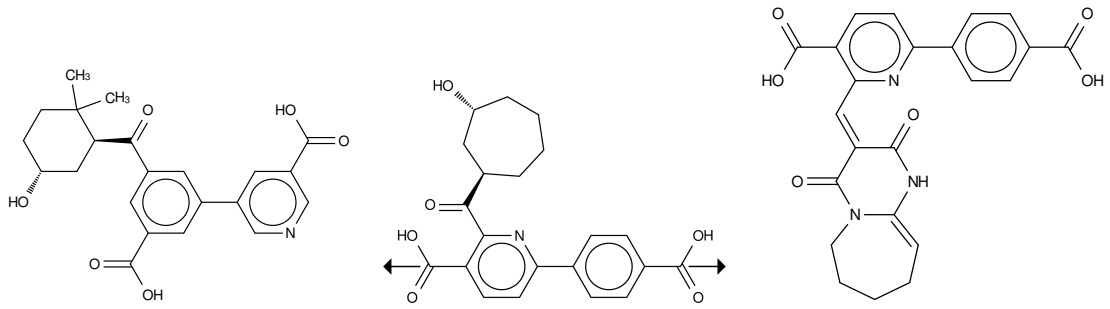
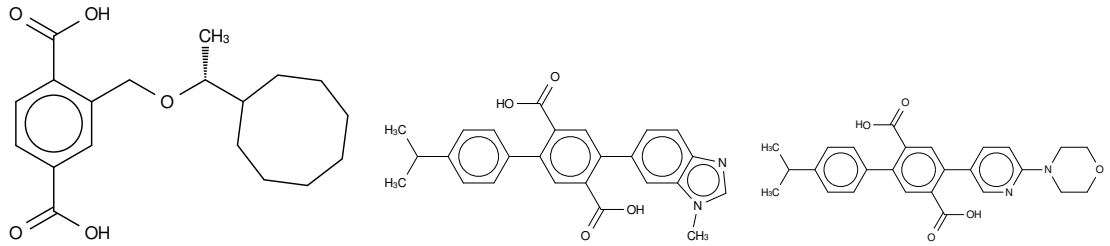


Figure 5.2 : Distributions of linker molecules in GSA–VSA space. (a) GSA and VSA of evolved linkers for final populations in each of the nine MOF nets and two precursor libraries. (b) The Pareto optimal frontiers for each MOF net of precursor library A. (c) The Pareto optimal frontiers for each MOF net of precursor library B. (d) The Pareto optimal frontiers for each MOF net of both libraries. In all cases, only molecules with pairwise angle $\geq 155°$ are shown.



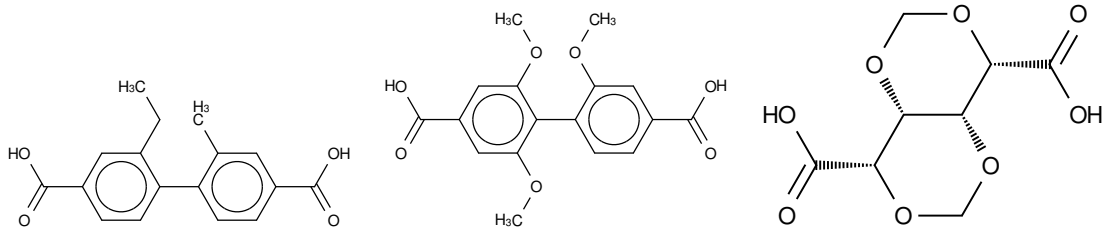
(a) acs-A-13826100 (b) acs-B-13955300 (c) acs-B-14132400

Figure 5.3 : Representative linker molecules of **acs** net. Each molecule is named as net-library-$GSA \cdot VSA$. Each black arrow in (b) is a vector orienting from the carbon atom and passing through the midpoint of two oxygen atoms for the carboxylic group metal attachment site. The pairwise angle of the linker molecule is defined as that between these two vectors. We require the pairwise angle $\geq 155°$ or $\geq 160°$.



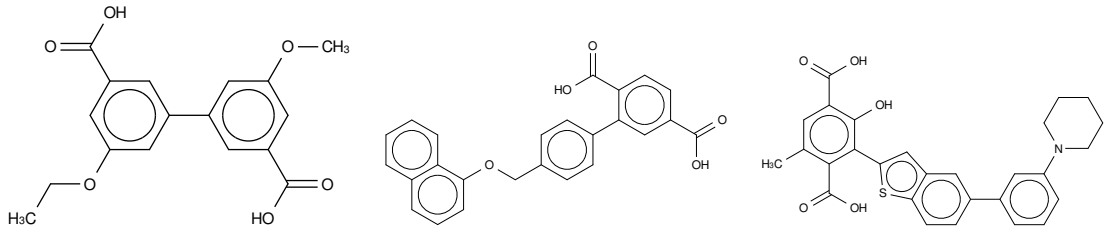
(a) dia-A-14053700 (b) dia-B-18266400 (c) dia-B-18870200

Figure 5.4 : Representative molecules of **dia** net.



(a) pcu-A-11797800 (b) pcu-A-12471900 (c) pcu-B-10469300

Figure 5.5 : Representative molecules of **pcu** net.



(a) sod-A-7820700 (b) sod-B-16415500 (c) sod-B-17885400

Figure 5.6 : Representative molecules of **sod** net.

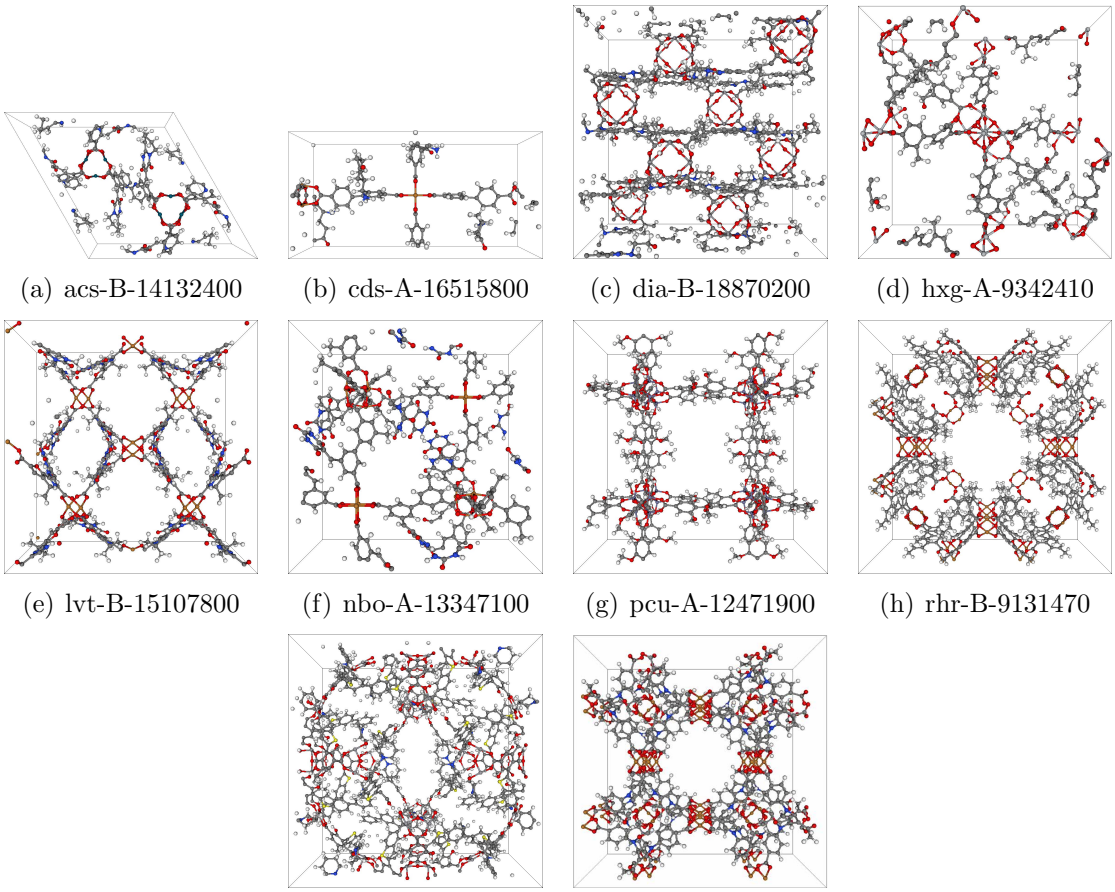
The deliverable capacities at the 65–5.8 bar and 35–5.8 bar conditions for the results using precursor library B are shown in Fig. 5.9(a)–(d).

5.4 Results of Optimizing Deliverable Capacity at 65–5.8 bar Condition

MOFs were evolved to optimize the methane deliverable capacity at 65–5.8 bar and 298 K. This set of results contains 3035 MOFs, and only precursor library B is used. The deliverable capacity at the 35–5.8 bar condition and $GSA \cdot VSA$ are also calculated for each MOF, as summarized in Table. 5.3 and plotted in Fig. 5.10.

5.5 Results of Optimizing Deliverable Capacity at 35–5.8 bar Condition

MOFs were evolved to optimize the methane deliverable capacity at 35–5.8 bar and 298 K. This set of results contains 4141 MOFs (3820 MOFs for the nine nets and 321 MOFs for **tbo** net), and only precursor library B is used. The deliverable capacity at the 65–5.8 bar condition and $GSA \cdot VSA$ are also calculated for each MOF, as summarized in Table. 5.4 and plotted in Fig. 5.11.



(a) acs-B-14132400 (b) cds-A-16515800 (c) dia-B-18870200 (d) hxg-A-9342410

(e) lvt-B-15107800 (f) nbo-A-13347100 (g) pcu-A-12471900 (h) rhr-B-9131470

(i) sod-B-17885400 (j) tbo-B-8794350

Figure 5.7 : Representative MOF conformations in each of ten nets. Each conformation shows the geometry of the MOF for a given linker and SBUs, as noted in Table 5.1 and Fig. 5.1. tbo-B-8794350 is from the set of deliverable capacity at 35–5.8 bar condition optimized. The other nine conformations are from the set of $GSA \cdot VSA$ optimized.

At the 35–5.8 bar condition, HKUST-1, a **tbo** net, has been a well-tested MOF with a deliverable capacity of 141 v(STP)/v experimentally [79]. To compare our results with HKUST-1, we perform the algorithm on the **tbo** net in addition to the nine nets at the 35–5.8 bar condition. Taking the requirement of three-connecting linkers into account, we customize the filters for the **tbo** net as discussed previously. The distribution of deliverable capacities of MOFs in ten nets is shown in Fig. 5.12.

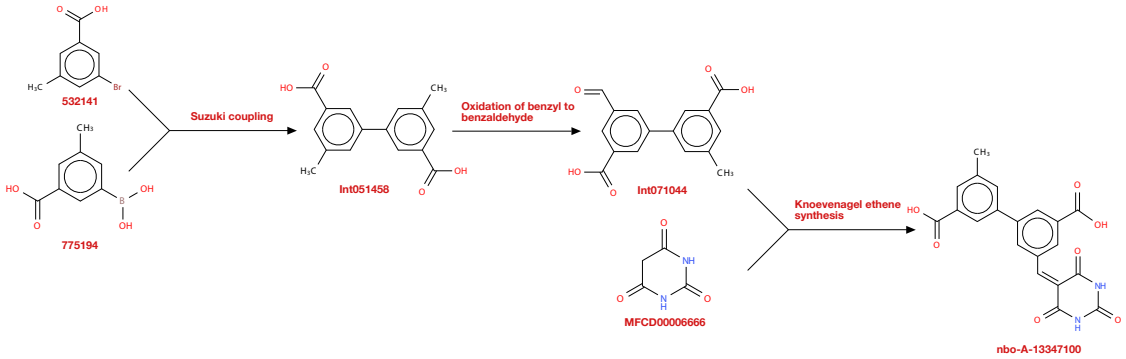


Figure 5.8 : Synthetic route of molecule nbo-A-13347100 in **nbo** network produced from precursor library A. Compounds 532141 and 775194 from the precursor library are Suzuki coupled to produce the intermediate product Int051458. The benzyl of Int051458 is oxidized into benzaldehyde, and Int071044 is formed. Int071044 and compound MFCD00006666 from the precursor library undergo Knoevenagel ethene synthesis and produce the final product nbo-A-13347100.

Net	Number of Candidates	Highest D.C. at the 65 bar condition	D.C. at the 35 bar condition	$GSA \cdot VSA$	Pearson Correlation
acs	315	183.62	128.45	8,601,250	0.826
cds	409	197.87	143.62	12,090,800	0.925
dia	150	177.32	144.77	13,471,500	0.209
hxg	516	171.99	115.67	2,840,570	0.457
lvt	935	189.53	127.95	296,427	0.318
nbo	374	182.08	121.28	6,728,520	0.580
pcu	201	186.22	132.66	1,964,840	0.748
rhr	42	158.75	99.60	3,739,120	-0.486
sod	93	153.93	97.51	3,230,480	0.301

Table 5.3 : Summary of MOFs with top deliverable capacity at the 65 bar condition for each net. Only results from linkers with pairwise angle $\geq 160°$ are shown. The best predicted MOF at the 65–5.8 bar condition is a **cds** net with a deliverable capacity of 197.87 v(STP)/v. Deliverable capacity at the 35–5.8 bar condition and $GSA{\cdot}VSA$ are shown for the best optimized MOF in each net. The Pearson correlation coefficient is calculated between highest deliverable capacity at the 65 bar condition and $GSA{\cdot}VSA$ (column 3 and column 5) for each net, and heavily depends on the net. The average Pearson correlation coefficient for all nine nets is 0.481. High deliverable capacity at the 65 bar condition does not necessarily indicate high $GSA \cdot VSA$. Deliverable capacity: v(STP)/v; $GSA \cdot VSA$: $m^4/(g\ cm^3)$

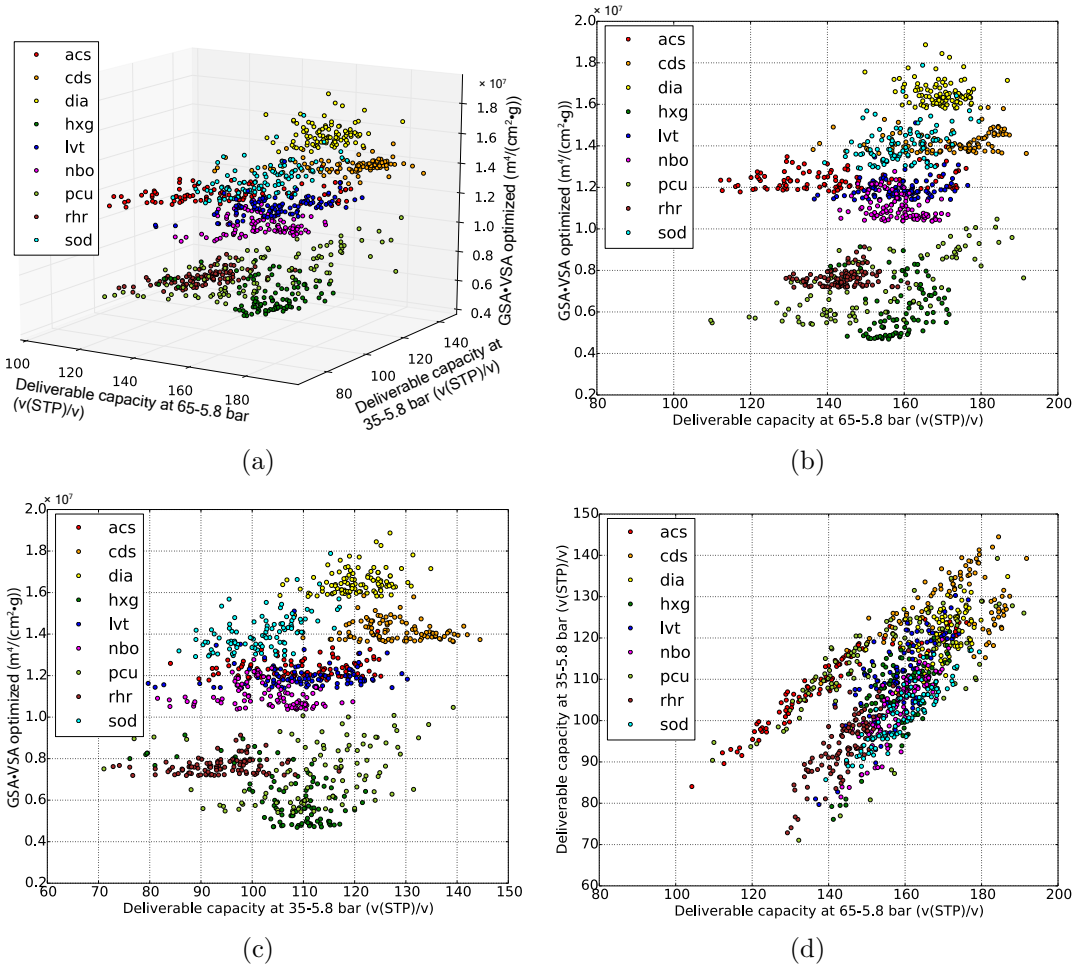


Figure 5.9 : Relationship between optimized surface area and deliverable capacities using precursor library B. (a) The points represent the surface area optimized results in the space of $GSA \cdot VSA$ and deliverable capacities. (b,c) MOFs with a certain $GSA \cdot VSA$ can have a range of deliverable capacities, although high $GSA \cdot VSA$ tend to indicate high deliverable capacities on average. (d) Deliverable capacities at two pressure conditions have a positive correlation.

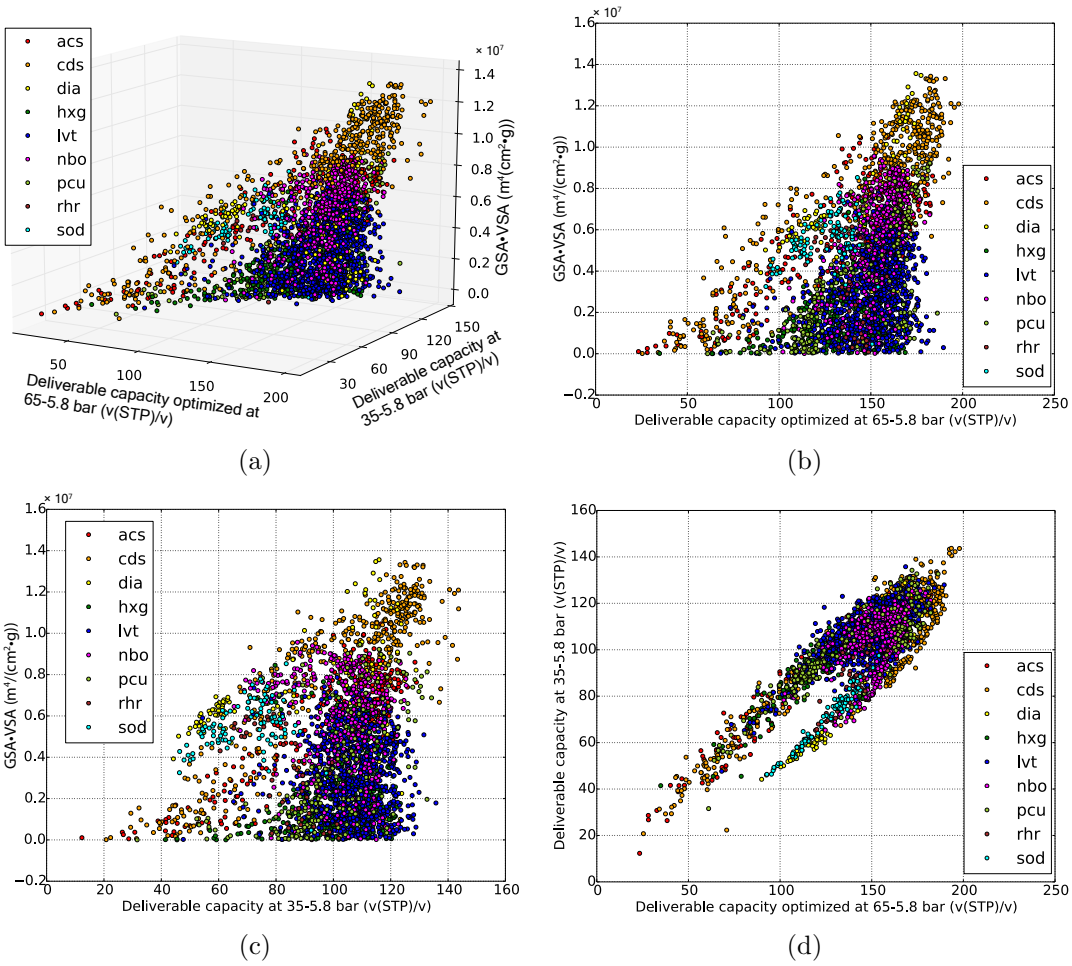


Figure 5.10 : Relationship between optimized deliverable capacity at the 65 bar condition, deliverable capacity at the 35 bar condition and surface area. (a) **cds** tends to occupy the highest end in the 3-dimensional phase space, but has a broad distribution. A large number of MOFs in **nbo** and **lvt** are located in a similar concentrated region, while other nets, for example **acs** and **sod**, exhibit a long-band pattern. (b, c) For a certain deliverable capacity, the distribution of $GSA \cdot VSA$ strongly depends on network. For instance, high deliverable capacities tend to indicate high $GSA \cdot VSA$ in **cds**; while $GSA \cdot VSA$ is almost independent from deliverable capacities in **nbo**. (d) Deliverable capacities at the two pressure conditions have positive correlations.

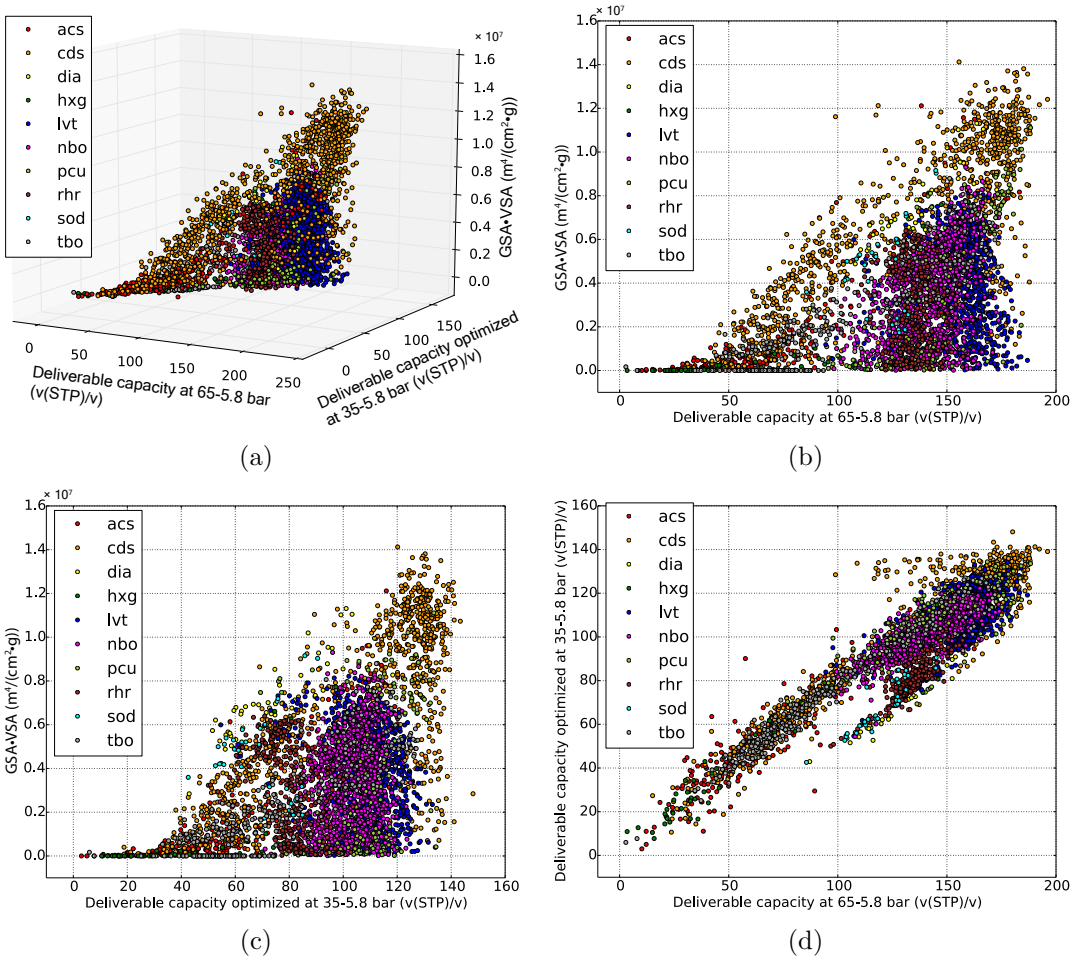


Figure 5.11 : Relationship between optimized deliverable capacity at the 35 bar condition, deliverable capacity at the 65 bar condition, and surface area. (a, b, c, d) The distributions show similar patterns to the results of optimized deliverable capacity at the 65–5.8 bar condition.

Net	Number of Candidates	Highest D.C. at the 35 bar condition	D.C. at the 65 bar condition	$GSA \cdot VSA$	Pearson Correlation
acs	259	123.72	157.57	11,574,200	0.872
cds	926	148.09	180.12	2,844,200	0.778
dia	45	118.34	172.82	8,271,250	0.536
hxg	84	121.03	160.46	3,228,680	0.678
lvt	848	134.11	177.76	396,580	-0.220
nbo	1,223	124.89	167.40	3,351,830	0.228
pcu	122	134.37	168.52	417,578	0.406
rhr	280	103.62	149.76	2,247,810	-0.357
sod	33	90.51	145.08	9,031,330	-0.049
tbo	321	127.62	170.02	5,288,910	0.860

Table 5.4 : Summary of MOFs with top deliverable capacity at the 35 bar condition for each net including **tbo** net. Only results from linkers with pairwise angle $\geq$ 160° are shown. The best predicted MOF at the 35–5.8 bar condition is a **cds** net with a deliverable capacity of 148.09 v(STP)/v. Deliverable capacity at the 65–5.8 bar condition and $GSA \cdot VSA$ are shown for The Pearson correlation coefficient is calculated between highest deliverable capacity at the 35 bar condition and $GSA \cdot VSA$ (column 3 and column 5) for each net, and heavily depends on the net. The average Pearson correlation coefficient for all nine nets is 0.532. High deliverable capacity at the 35 bar condition does not necessarily indicate high $GSA \cdot VSA$. Deliverable capacity: v(STP)/v; $GSA \cdot VSA$: $m^4/(g\ cm^3)$

We find 319 MOFs in the nets of **cds**, **lvt**, and **pcu**, and 2 MOFs in the net of **tbo** with a higher deliverable capacity than HKUST-1.

5.6 Discussion and Conclusion

Achievable surface areas for nine of these MOF networks have previously been reported [92], where shapes of potential MOF linkers in an abstract chemical space were optimized for surface area. The results shown here are exciting because they validate the high surface areas that were previously predicted using shape only, and are now exemplified with molecular structure. Interestingly, the evolution procedure developed here to identify chemically-synthesizable linkers leads to surface area val-

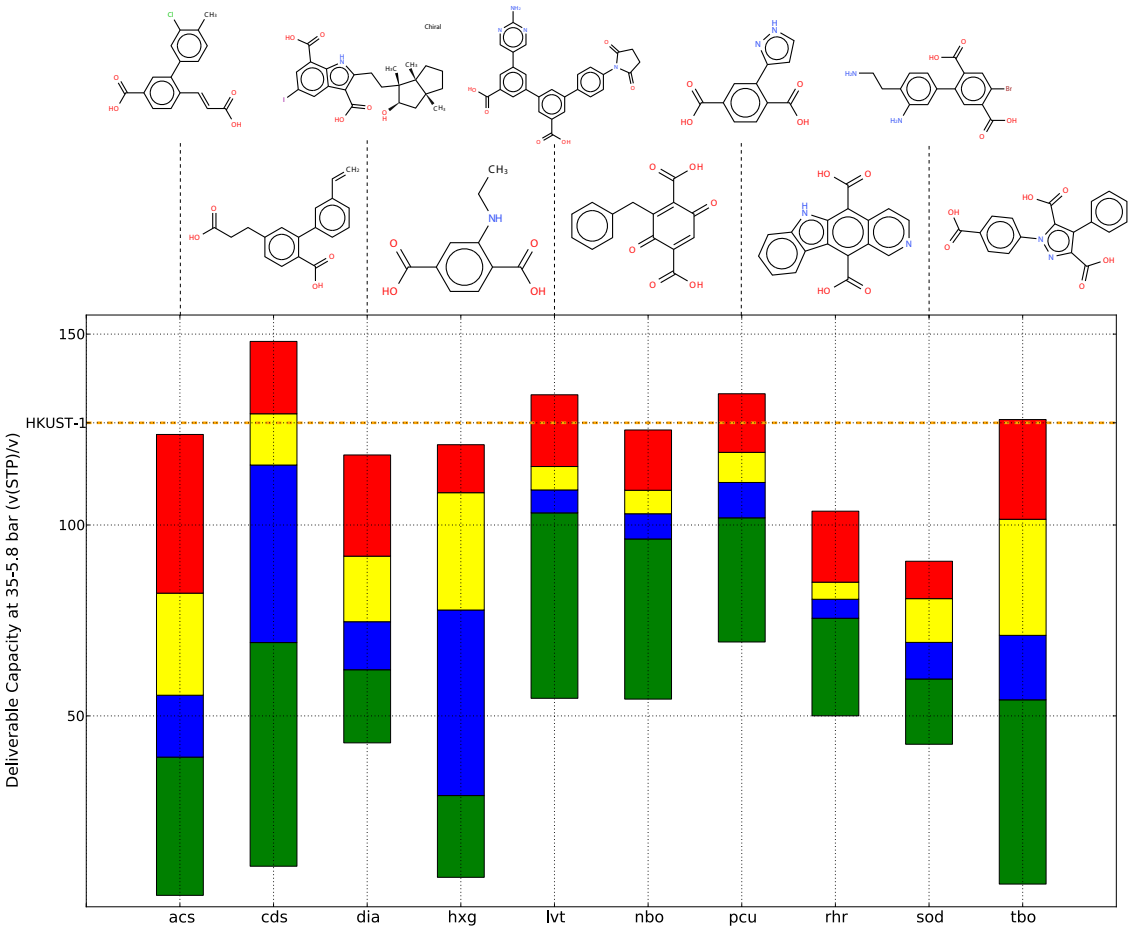


Figure 5.12 : Distribution of deliverable capacity at the 35–5.8 bar condition of all MOFs identified in the ten nets. Only results from linkers with pairwise angle $\geq 160°$ are shown. HKUST-1 is found in the results of **tbo** net with a computed deliverable capacity of 126.77 v(STP)/v. The best organic linker of each net is shown above each bar. Each color of the bars represents one quartile of the results.

ues, $GSA \cdot VSA$, superior than those previously reported for optimized linker shapes for the nets **acs**, **dia**, **lvt**, **nbo**, **pcu**, **rhr** and **sod**, as shown in Table 5.2.

The Pareto frontiers of different nets in Fig. 5.2(d) shows that the **lvt** net dominates the **nbo** net, i.e. the maximal GSA, VSA, and $GSA \cdot VSA$ of **lvt** are bigger than those of **nbo**, while the **pcu** net dominates the **hxg** net. Furthermore, **dia** dominates many nets, as does **sod**. Fig. 5.2(a) provides even greater insight, particularly into the comparative behaviors of the precursor libraries. Clearly, the **sod** libraries have radically different performance. A less drastic differentiation may also be drawn for the other nets. For instance, the MOFs produced from the different libraries occupy different regions of the property space. MOFs in the **pcu** net produced from precursor library A are consistently in one central region, while those produced from precursor library B are just outside this region and mainly extend in the decreasing GSA direction as well as explore a few outlying positions.

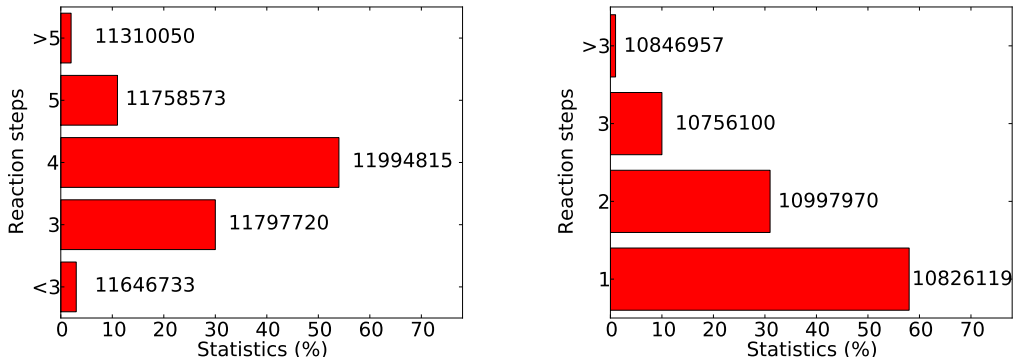
Since the procedure is stochastic, different runs on the same net and with the same precursor library can lead to different final populations of linkers and different optimized values. Nevertheless, the tendency to improve optimizing values as the generation increases appears general [32]. A set of strict selection criteria tends to lead to small populations of elite compounds, while a set of less strict selection criteria leads to a bigger and more diverse population of compounds. For this reason, in the set of surface area optimized results, some of the runs use a looser angle filter to facilitate exploration of the linker chemical space. For example, an angle filter of $\geq 125^\circ$ is used for the networks **dia**, **sod**, and **lvt** with precursor library A. This filter may not be necessary if a strict filter could lead to many good candidates. A tighter criterion should be applied if too many mediocre candidates are generated. This is the case with all the networks when using precursor library B, for which we

use an angle filter of $\geq 155°$. It is worth noting that the mean pairwise angle of linkers cannot be strictly 180° because the conformation of molecules fluctuates at finite temperature. For instance, the organic linker BPDC in MOF-106, a known "linear" linker, is observed in a bent conformation, which leads to corrugated sheet structures [95].

Several of the identified linkers have long side groups that could be important to increase the surface area. Generally, high GSA linkers tend to be long or branched, so as to maximize surface area with as much light-weight organic material as possible. Conversely, high VSA is usually the result of utilizing short linkers, so as to minimize the volume of material and increase the surface area by growing linker width. The product of GSA and VSA considers these two competing effects [92] and has been used in study of hydrogen storage potential [93]. Intuitively, we can anticipate that the highest $GSA \cdot VSA$ will be the result of short organic linkers, to minimize material volume, with significant branching, to maximize the contribution of the organic to the surface area.

We constrain the total number of reaction steps for any of our predicted linkers to ten. Nevertheless, it is rare to identify predicted linkers that are synthesized with five or more than five steps in our results. Fig. 5.13 shows the statistics of different numbers of reaction steps in the final population in the **dia** and **pcu** nets with precursor library A for the set of surface area optimized results. The **dia** net significantly favors more reaction steps than does the **pcu** net. We ascribe this distinction to two reasons: 1) the nature of a network determines how difficult it is to grow a complicated molecule within the network; and 2) values of $GSA \cdot VSA$ may set an upper limit in which molecules with only a few reaction steps may already be sufficiently outstanding, as is probably the case for the **pcu** net. Predicted molecules

with too many synthesis steps will not be attractive as practical candidates because a long synthetic route leads to low yield and high cost of synthesis [73]. We note that mean $GSA \cdot VSA$ for the linkers with different number of reaction steps are approximately equal, as shown in Fig. 5.13. This result suggests that high $GSA \cdot VSA$ can be achieved with relatively few reaction steps.



(a) **dia** net (b) **pcu** net

Figure 5.13 : Statistics of observing linkers with different numbers of reaction steps in the final populations of MOFs optimized for surface area, for two nets with precursor library A. The numbers to the right side of the bars are mean $GSA \cdot VSA$ for each number of reaction steps.

We used two precursor libraries of commercially available compounds in this study. This was done not only to provide statistics but also to judge which precursor library might lead to better performance. Overall, both precursor libraries A and B lead to similar results, except for the **dia** and **sod** network, in which precursor library B achieves higher values, as shown in Table 5.2. Precursor library A contains specific molecules with one carboxylic group and one benzyl group, while precursor library B is a balanced library, containing medicinal chemistry compounds and organic building blocks from Sigma-Aldrich. Library B is a fairly universal one for use under general conditions. Therefore, we used precursor library B for the sets of deliverable capacity

optimized results. It is noteworthy that besides library difference, **dia** and **sod** used an angle filter of $\geq 125^\circ$ in precursor library A and $\geq 155^\circ$ in precursor library B. Using an angle filter of 155° in precursor library A leads to $GSA \cdot VSA$ values of 13,270,500 $m^4/(g\ cm^3)$ for **dia** and 13,650,400 $m^4/(g\ cm^3)$ for **sod**, which suggests that the effect of the angle filter on the evolution of the population of linkers is network dependent.

Using $GSA \cdot VSA$ and deliverable capacities as metrics for surface area and gas adsorption performance respectively, we find that the correlation between them in the optimized materials presented here strongly depends on the network. For example, **rhr** in Table 5.3, and **lvt**, **rhr** and **sod** in Table 5.4 show a negative correlation of surface area and deliverable capacity. Thus, caution must be exercised when surface area of MOFs is used to forecast potential adsorption capacity. The deliverable capacities at the 65–5.8 bar and 35–5.8 bar conditions are strongly positively correlated as shown in Fig. 5.9(d), Fig. 5.10(d) and Fig. 5.11(d), because these two quantities are derived from the same isothermal curve at 298 K. A similar relation between these two deliverable capacities was recently observed for covalent organic frameworks (COFs) [96], and porous polymer networks (PPNs) [97].

One intriguing perspective is to see whether experimentally synthesized MOFs, which have outstanding deliverable capacity, have high surface area. MOF-5, a **pcu** net, has been one benchmark MOF with a reported high methane deliverable capacity [49, 78, 79]. It is found in our results with a $GSA \cdot VSA$ of 10,007,100 $m^4/(g\ cm^3)$, which is not among the highest. For instance, within the set of $GSA \cdot VSA$ optimized results, 103 MOFs of the **pcu** net have superior $GSA \cdot VSA$ values, and three are shown in Fig. 5.5(a)–(c) with the highest $GSA \cdot VSA$ of 12,471,900 $m^4/(g\ cm^3)$. In total, 1,206 results have superior $GSA \cdot VSA$ values than MOF-5 in all nine nets.

HKUST-1 is a **tbo** net with a high deliverable capacity, and it is found in the set of predicted results when deliverable capacity is optimized at the 35–5.8 bar condition. This MOF has a $GSA{\cdot}VSA$ of 4,914,780 $m^4/(g\ cm^3)$, compared to which 56 predicted MOFs of the **tbo** net have higher values.

Synthesizing these predicted MOFs requires consideration of several additional practical matters. For example the solubility of a proposed linker in a specific solvent and at a synthesis temperature is an important variable. Solubility is easily estimated, and it may be included in the scoring function of the evolutionary algorithm presented here. Alternatively, since multiple solvents and co-solvents are available for MOF synthesis, they may be chosen to accommodate a particular self-assembly of predicted linker and SBU nicely.

In summary, many chemically synthesizable linkers, which have considerable rigidity, suitable metal node binding sites, and promising $GSA{\cdot}VSA$ values, were identified in nine MOF networks. The **dia** network appears to be an excellent choice for compromising $GSA \cdot VSA$ surface area, with **cds** and **sod** also being good choices. The networks **lvt**, **cds**, and **sod** stand out for GSA alone, while **dia**, **cds**, and **acs** stand out for VSA alone.

Two sets of MOFs with optimized deliverable capacities at the 65–5.8 bar condition and the 35–5.8 bar condition were also identified. MOFs with high deliverable capacities are not necessary to have high $GSA \cdot VSA$. Conversely, MOFs with high $GSA \cdot VSA$ tend to have high deliverable capacities on average, but exceptions occur for specific networks, and negative correlations between $GSA \cdot VSA$ and deliverable capacities are observed. We notice that the highest deliverable capacity at 65–5.8 bar condition obtained in this work is 197.87 v(STP)/v, which agrees with a recent proposed upper limit of approximately 200 v(STP)/v [2]. Some structures from the

combinatorially created database approach this limit [17], and structures produced here approach that limit as well. The distribution of deliverable capacities of the structures produced with the present method is centered closer to this limit than are those of the combinatorial database [2]. In other words, optimization of deliverable capacity by the present method produces a larger set of MOFs with high deliverable capacity than does screening of a combinatorially created database.

Chapter 6

Flexibility of MOFs

6.1 Overview

Porous MOFs have a crystal structure with open space for hosting guest molecules, and this feature is key to gas adsorption [98]. Porosity can endow crystalline networks with high susceptibility to environmental changes, for example, tempertaure, mechanical stress or gaseous pressure. Thermal expansion is a significant factor even in tranditionally "rigid" MOFs, such as MOF-5 [49, 99] and HKUST-1 [100]. However, the flexibility I are talking about here is a more dramatic change of unit cell volume and shape. This flexibility sometimes termed as "breathing effect" or "gate opening" can lead to a phase transition of the crystal structures. The phase transition occurs at a characteristic temperature, stress or pressure.

In 2014, less than 0.5% of the MOFs showed a phase transition in the Cambridge Structural Database (CSD). The progress, mechanisms, and applications of flexible MOFs have been reviewed [45]. The early MIL-53 MOF variants show an extraordinary breathing effect by closing and reopening the channels [101, 43, 44, 102]. Other important flexible MOFs include MIL-88 [103], ZIF-8 [104], and Fe(bdp) and Co(bdp) analogs [46, 47] for which more discussion will be given in Section 6.2.

The phase transitions of these MOFs have been termed as large pore (LP) phase vs. narrow pore (NP) phase, high temperature (HT) phase vs. low temperature (LT) phase, or expanded phase vs. collapsed phase. These terms are used according to the

context of particular research.

6.2 Heat of Adsorption

The gas adsorption process is accompanied by a heat transfer between the environment and the studied system. Thermodynamically the heat of adsorption comes from the free energy difference of guest molecules between the gas phase and the adsorbed phase. Experimentally the heat of adsorption may be obtained by calorimetric measurements. Computational simulations may be performed to calculate, for example, "isosteric heat" for heat of gas adsorption in a rigid framework [105]. The total heat of adsorption is the integral of isosteric heat from a given starting pressure or loading to an ending pressure or loading. Reviews on thermodynamics of gas adsorption in general host–guest system have been reported [106, 107].

We are interested in heat of adsorption because of the recent progress in flexible MOF synthesis. Long's group synthesized a new group of MOFs, Fe(bdp) and Co(bdp) analogs [46, 47], which not only show an interesting structural phase transition of the framework between an expanded phase at high methane pressure and a collapsed phase at low methane pressure, but also exhibit a beneficial heat management due to the transition. The phase transition results in an "S-shaped" methane isotherm, which potentially increases the deliverable capacity for 35 bar adsorption pressure and 5.8 bar delivery pressure, compared with a classical type-I isotherm [48]. The phase transition can reduce the heat of adsorption that is a significant practical but frequently overlooked problem in gas storage material development. In sorption, the energy needed to overcome the barrier from the collapsed phase to the expanded

phase is compensated by the enthalpy of methane adsorption, as shown in Eq. 6.1.

$$Q = Q_{ads} + Q_{ph.tran.} = -\Delta H_{ads} + \Delta H_{ph.tran.} \tag{6.1}$$

Thus, this phase transition feature provides the Fe(bdp) and Co(bdp) analogs with internal heat management in the adsorption and desorption processes. These intriguing discoveries inspire us to study potential flexible MOFs that undergo a structural phase transition and have a lower heat of adsorption and greater gas deliverable capacity.

To evaluate heat of adsorption, researchers have developed theories for calculating the so-called "isosteric heat" for rigid frameworks [108, 109]. As the word "isosteric" suggests, the spatial arrangement of adsorbent molecules is assumed to be unchanged in adsorption. One method to calculate isosteric heat is to start from Clausius-Clapeyron equation which gives

$$\begin{aligned} H_{st}(N) &= RT^2 \left(\frac{\partial \ln P}{\partial T} \right) \Big|_N \\ &\approx \frac{R \ln \left(\frac{P^*}{P} \right)}{\frac{1}{T^*} - \frac{1}{T}} \Big|_N \end{aligned} \tag{6.2}$$

where isosteric heat is given as a function of loading N. Notice that T and P are temperature and pressure at which I simulate, T^* and P^* are temperature and pressure at which the loading has a same value of N. For a given isotherm of temperature T, I may choose another reference isotherm of temperature T^*. A given loading N at pressure P in the first isotherm corresponds to a same loading at pressure P^* in the second isotherm.

In Fig. 6.1, I present isosteric heat results of simulating methane adsorption in

MIL-53(Al)ht where "ht" means high temperature phase. The GCMC simulation is

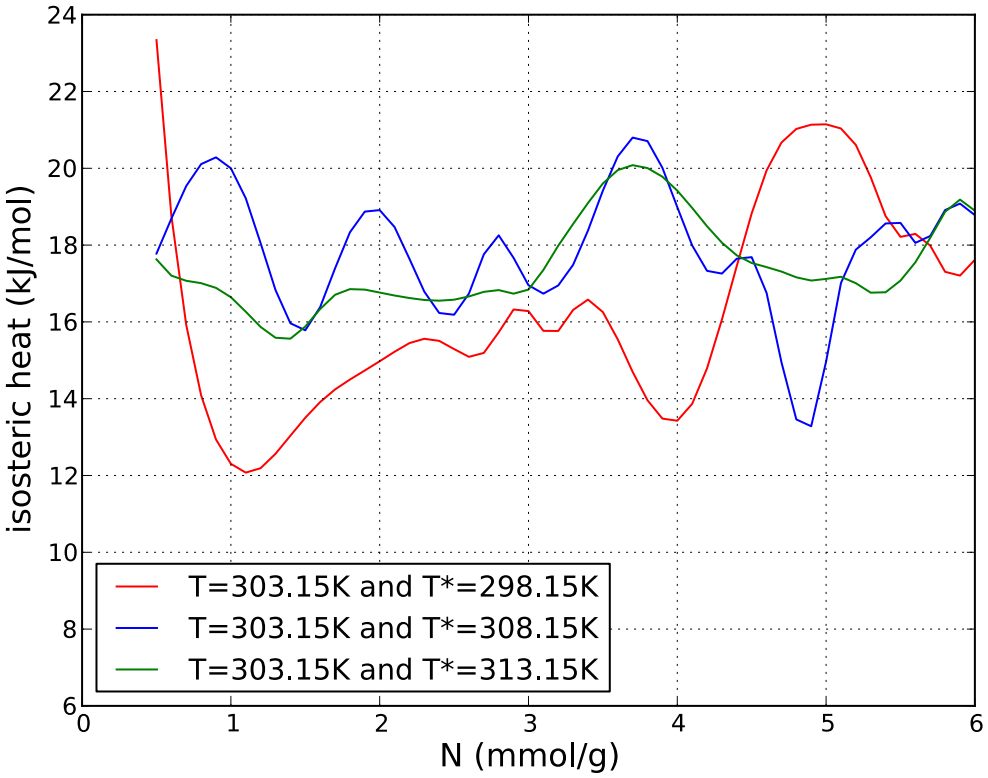


Figure 6.1 : Isosteric heat of rigid MIL-53(Al)ht loading methane using Clausius-Clapeyron method. Three reference temperatures are chosen as 298.15K, 308.15K, or 313.15K. The UFF force field is used to simulate the underlying isotherm. Curves are fitted to the simulation results.

performed in RASPA [76], and the framework is treated as rigid. Notice that isosteric heat is a differential quantity, and heat is path-dependent. An integral of isosteric heat over a range of independent thermodynamic variable gives the desired heat of adsorption.

A second method to calculate isosteric heat is the so-called fluctuation method. More discussion of this form will be given in Section 6.3. The formula of molar

isosteric heat is given by

$$q_{st} = k_B T + u_g - \frac{\langle UN \rangle - \langle U \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \tag{6.3}$$

where u_g is the molar internal energy of gas molecules in the ideal gas phase. The brackets $\langle\rangle$ denote ensemble average of a certain quantity. The U is the internal energy of both framework and gas molecules in the adsorbed phase, and N is the number of gas molecules in the adsorbed phase. In the context of rigid frameworks, the grand canonical ensemble is traditionally used in gas adsorption simulations.

In Fig. 6.2, I present isosteric heat results of methane adsorption in MIL-53(Al)ht using the fluctuation method.

Fig. 6.1 and Fig. 6.2 may be compared with Figure 7 of a previous research [110] with qualitative agreement. A closer look tells us that the results of the fluctuation method show an upward trend as gas uptake goes up, which better matches experimental results. Therefore, I am going to discuss the extension of the fluctuation method to flexible frameworks in the next section.

6.3 Heat of Adsorption in a Flexible Framework

The original form of the isosteric heat, assuming the gas phase is ideal, is

$$H_{st} = RT^2 \left(\frac{\partial \ln P}{\partial T} \right) \Bigg|_N \tag{6.4}$$

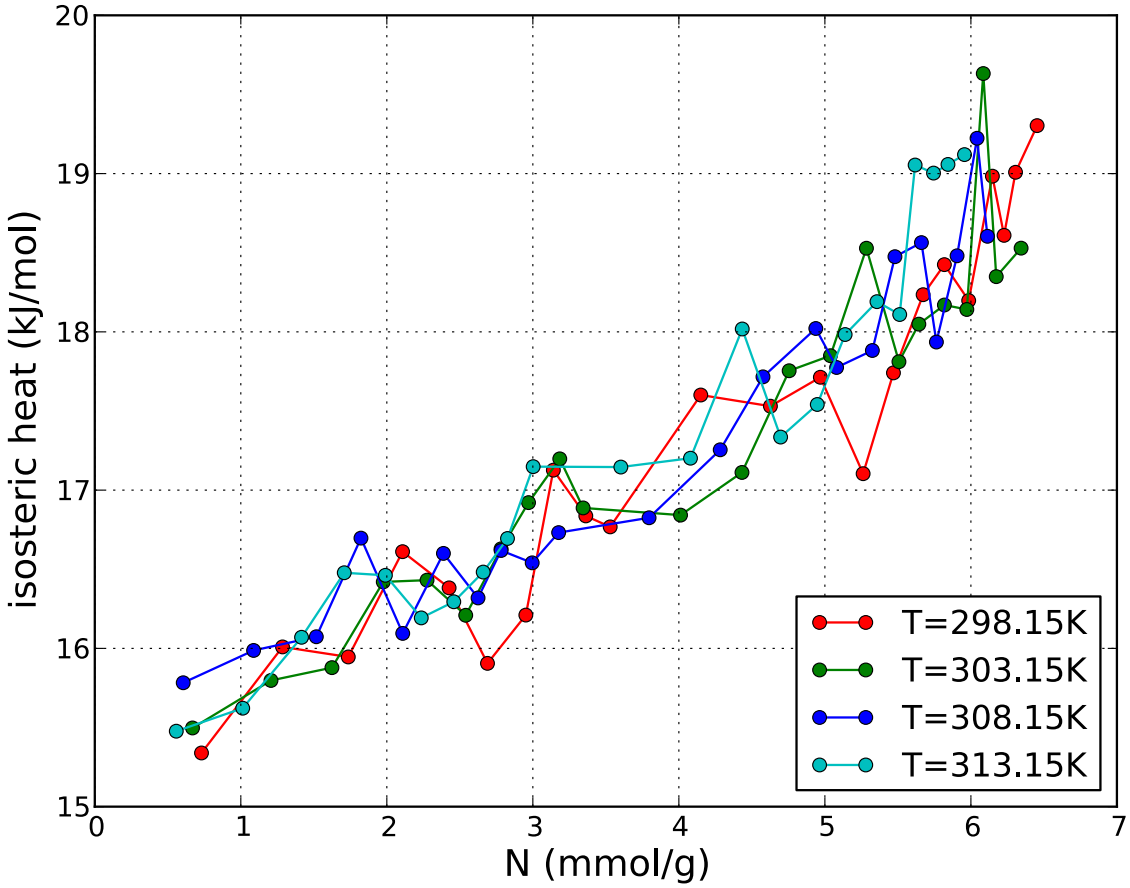


Figure 6.2 : Isosteric heat of rigid MIL-53(Al)ht loading methane using the fluctuation method. The four curves are derived from isotherms at four different temperatures. The UFF force field is used to simulate the system.

where R is gas constant, T is temperature, P is pressure, and N is number of gas molecules. In grand canonical ensemble (μVT ensemble), the molar isosteric heat is

$$q_{st} = k_B T + u_g - \frac{\langle UN \rangle - \langle U \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \tag{6.5}$$

where u_g is the molar internal energy of the ideal gas, and U is internal energy of the system in the adsorbed phase.

For a flexible framework, I consider the volume of system to be changeable. The volume depends on pressure and temperature. The number of framework atoms is fixed. Therefore I need to choose an ensemble $(\mu(P), N_{MOF}, P, T)$, where μ is chemical potential. To derive the expressions of thermodynamics quantities in the $(\mu(P), N_{MOF}, P, T)$ ensemble, I need to know the partition function. A starting point is the partition function of (N, N_{MOF}, P, T) ensemble with two species [38]

$$\begin{aligned} &\mathscr{Q}'(N, N_{MOF}, P, T) \\ =&\frac{\beta P}{\Lambda^{3N} \Lambda_{MOF}^{3N_{MOF}} N! N_{MOF}!} \int dV V^{N+N_{MOF}} \exp(-\beta P V) \int d\vec{s}^{N+N_{MOF}} \exp[-\beta E_{tot}(\vec{s}^{N+N_{MOF}})]. \end{aligned} \tag{6.6}$$

Eq. 6.6 can be converted to a partition function for $(\mu(P), N_{MOF}, P, T)$ ensemble by summing over N,

$$\mathscr{Q}(\mu(P), N_{MOF}, P, T) = \sum_N \exp(-\alpha N) \mathscr{Q}'(N, N_{MOF}, P, T) \tag{6.7}$$

where $\alpha = -\frac{\mu}{k_B T}$.

The heat flow for the system to change from state 1 to state 2 at fixed N_{MOF} and

T is given by

$$\begin{aligned} Q =& E_{tot}(\mu(P_2), N_{MOF}, P_2, T) - E_{tot}(\mu(P_1), N_{MOF}, P_1, T) - E_{env}(\Delta N, P_1, T) \\ &+ \int_{P_1}^{P_2} P_{ext} \frac{dV}{dP_{int}} dP_{int} - \int_{P_1}^{P_2} P_{int} v_{env} \frac{dN}{dP_{int}} dP_{int} \end{aligned} \tag{6.8}$$

where E_{tot} is the total energy of system, E_{env} is the energy of gas molecules in the external environment before joining the system, v_{env} is the unit volume of gas molecules in the environment, P_{int} is the internal pressure of the system, and P_{ext} is the external pressure of environment. The last two terms represent the work and $P_{int} = P_{ext}$ at equilibrium. At a given state, E_{tot} fluctuates over time. The first two terms can be expressed in terms of ensemble averages

$$\begin{aligned} &\langle E_{tot} \rangle(\mu(P_2), N_{MOF}, P_2, T) - \langle E_{tot} \rangle(\mu(P_1), N_{MOF}, P_1, T) \\ \approx& \int_{\mu_1}^{\mu_2} \left. \frac{\partial \langle E_{tot} \rangle}{\partial \mu} \right|_P d\mu + \int_{P_1}^{P_2} \left. \frac{\partial \langle E_{tot} \rangle}{\partial P} \right|_\mu dP \\ =& \int_{P_1}^{P_2} \left. \frac{\partial \langle E_{tot} \rangle}{\partial \alpha} \right|_P \left. \frac{\partial \alpha}{\partial \mu} \right|_T \left. \frac{\partial \mu}{\partial P} \right|_T dP + \int_{P_1}^{P_2} \left. \frac{\partial \langle E_{tot} \rangle}{\partial P} \right|_\mu dP \\ =& \int_{P_1}^{P_2} \beta \left(\langle E_{tot} N \rangle - \langle E_{tot} \rangle \langle N \rangle \right) \left|_P \left. \frac{\partial \mu}{\partial P} \right|_T dP + \int_{P_1}^{P_2} \left(\langle E_{tot}(-\beta V) \rangle - \langle E_{tot} \rangle \langle -\beta V \rangle \right) \right|_\mu dP \\ =& \int_{P_1}^{P_2} \beta \left(\langle E_{tot} N \rangle - \langle E_{tot} \rangle \langle N \rangle \right) \left|_P v dP + \int_{P_1}^{P_2} \left(\langle E_{tot}(-\beta V) \rangle - \langle E_{tot} \rangle \langle -\beta V \rangle \right) \right|_\mu dP \end{aligned} \tag{6.9}$$

where v is the molar volume of gas molecules

$$\left. \frac{d\mu}{dP} \right|_T = v \tag{6.10}$$

and $\alpha = -\frac{\mu}{k_B T}$. Then Eq. 6.8 can be written as

$$Q = \int_{P_1}^{P_2} \beta \left(\langle E_{tot} N \rangle - \langle E_{tot} \rangle \langle N \rangle \right) \Big|_P v dP + \int_{P_1}^{P_2} \left(\langle E_{tot}(-\beta V) \rangle - \langle E_{tot} \rangle \langle -\beta V \rangle \right) \Big|_\mu dP$$
$$- \langle E_{env}(\Delta N, P_1, T) \rangle + \int_{P_1}^{P_2} P_{ext} \frac{d\langle V \rangle}{dP_{int}} dP_{int} - \int_{P_1}^{P_2} P_{int} v_{env} \frac{d\langle N \rangle}{dP_{int}} dP_{int} \tag{6.11}$$

where $v_{env} = \frac{V_{env}(\langle \Delta N \rangle, P_1, T)}{\langle \Delta N \rangle}$. The Eq. 6.11 gives a formula for calculating heat flow associated with transition from state 1 to state 2, but it is not convenient to be used in a simulation.

To derive a formula of isosteric heat that may be used in a computational simulation, I start from an equation in the Appendix of Woods *et al* [109]. The total entropy of the adsorbed fluid can be expanded as

$$T dS_a = dU_a - \mu_a dN_a + P dV_a \tag{6.12}$$

where the subscript a denotes the adsorbed phase of gas molecules. So

$$T \left(\frac{\partial S}{\partial N_a} \right)_T = \left(\frac{\partial U_a}{\partial N_a} \right)_T + P \left(\frac{\partial V_a}{\partial N_a} \right)_T - \mu_a. \tag{6.13}$$

Notice that the Gibbs free energy of the system can be expressed as

$$dG_a = -S_a dT + V_a dP + \mu_a dN_a. \tag{6.14}$$

Thus,

$$\begin{aligned} S_a &= -\left(\frac{\partial G_a}{\partial T}\right)_{P,N_a} \\ \mu_a &= \left(\frac{\partial G_a}{\partial N_a}\right)_{P,T}. \end{aligned} \tag{6.15}$$

Therefore,

$$\begin{aligned} \left(\frac{\partial S_a}{\partial N_a}\right)_T &= -\frac{\partial G_a}{\partial T \partial N_a} \\ \left(\frac{\partial \mu_a}{\partial T}\right)_{N_a} &= \frac{\partial G_a}{\partial N_a \partial T}. \end{aligned} \tag{6.16}$$

This Maxwell relationship allows me to rewrite Eq. 6.13 as

$$T\left(\frac{\partial \mu_a}{\partial T}\right)_{N_a} = \mu_a - \left(\frac{\partial U_a}{\partial N_a}\right)_T - P\left(\frac{\partial V_a}{\partial N_a}\right)_T. \tag{6.17}$$

At equilibrium,

$$d\mu_a = d\mu_g \tag{6.18}$$

where the subscript g denotes the gas phase. We consider $\mu_a = \mu_a(T, N_a)$ and $\mu_g = \mu_g(T, P)$. Thus

$$\left(\frac{\partial \mu_a}{\partial T}\right)_{N_a} dT = \left(\frac{\partial \mu_g}{\partial T}\right)_P dT + \left(\frac{\partial \mu_g}{\partial P}\right)_T dP. \tag{6.19}$$

The Maxwell relationship also holds for the gas phase, which gives

$$\left(\frac{\partial \mu_g}{\partial T}\right)_{N_g} = -\left(\frac{\partial S_g}{\partial N_g}\right)_T \tag{6.20}$$

and

$$\left(\frac{\partial \mu_g}{\partial P}\right)_{N_g} = \left(\frac{\partial V_g}{\partial N_g}\right)_P = v_g \tag{6.21}$$

where v_g is the molar volume of gas [111, 109]. Notice that

$$dS_g = \frac{1}{T}(dH_g - VdP - \mu_g dN_g). \tag{6.22}$$

With Eq. 6.22, I can rewrite the Eq. 6.20 as

$$\left(\frac{\partial \mu_g}{\partial T}\right)_P = \frac{\mu_g - h_g}{T} \tag{6.23}$$

where h_g is the molar enthalpy of gas. So, from Eq. 6.19,

$$\begin{aligned}\left(\frac{dP}{dT}\right)_{N_a} =& \frac{\left(\frac{\partial \mu_a}{\partial T}\right)_{N_a} - \left(\frac{\partial \mu_g}{\partial T}\right)_P}{\left(\frac{\partial \mu_g}{\partial P}\right)_T} \\ =& \frac{\frac{1}{T}(\mu_g - h_g) - \left(\frac{\partial \mu_g}{\partial T}\right)_{N_a}}{v_g}.\end{aligned} \tag{6.24}$$

Then according to Eq. 6.4, the isosteric heat is

$$\begin{aligned}H_{st} &= RT^2 \left(\frac{1}{P}\frac{\partial P}{\partial T}\right)_{N_a} \\ &= RT^2 \frac{1}{P}\left(\frac{dP}{dT}\right)_{N_a} \\ &= RT^2 \frac{1}{P}\frac{\frac{1}{T}\left[\mu_a - \left(\frac{\partial U_a}{\partial N_a}\right)_T - P\left(\frac{\partial V_a}{\partial N_a}\right)_T\right] - \frac{1}{T}(\mu_g - h_g)}{v_g} \\ &= \frac{RT}{P}\frac{\mu_a - \left(\frac{\partial U_a}{\partial N_a}\right)_T - P\left(\frac{\partial V_a}{\partial N_a}\right)_T - \mu_g + h_g}{v_g} \\ &= \frac{RT}{Pv_g}\left[h_g - \left(\frac{\partial U_a}{\partial N_a}\right)_T - P\left(\frac{\partial V_a}{\partial N_a}\right)_T\right]\end{aligned} \tag{6.25}$$

where I used $\left(\frac{dP}{dT}\right)_{N_a} = \left(\frac{\partial P}{\partial T}\right)_{N_a}$, and $\mu_a = \mu_g$. Thus, the molar isosteric heat is given

by

$$q_{st} = k_B T + u_g - \left(\frac{\partial U}{\partial N}\right)_T - P\left(\frac{\partial V}{\partial N}\right)_T. \tag{6.26}$$

It can be shown that

$$\begin{aligned} \frac{\partial \langle U \rangle}{\partial \langle N \rangle} &= \frac{\langle UN \rangle - \langle U \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \\ \frac{\partial \langle V \rangle}{\partial \langle N \rangle} &= \frac{\langle VN \rangle - \langle V \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2}. \end{aligned} \tag{6.27}$$

The proof needs calculations of $\frac{\partial \langle X \rangle}{\partial \alpha}$ for $X = U$, V, and N where $\alpha = -\frac{\mu}{k_B T}$. All the angle brackets $\langle\rangle$ are ensemble averages evaluated using the partition function in Eq. 6.7. Therefore, the molar isosteric heat of a flexible framework is derived from Eq. 6.26 and given by

$$q_{st} = k_B T + u_g - \frac{\langle (U + PV)N \rangle - \langle U + PV \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \tag{6.28}$$

Compared with Eq. 6.5, the only difference is that the internal energy U is changed into enthalpy, $H = U + PV$. This change is intuitive in the sense that the volume change of the system leads to a work done on the external environment.

Chapter 7

Simulation of Flexible MOFs

7.1 Overview

As stated in Section 6.2, we are inspired by the successful synthesis of Fe(bdp) and Co(bdp) MOF analogs [46, 47] to predict new flexible MOFs which potentially have superior heat management and higher deliverable capacity than ones having no phase transitions. Any breakthrough in this aspect might lead to a higher gas deliverable capacity limit [2]. In Chapter 4 and 5, I have shown that our evolutionary algorithm has great performance in searching the chemistry space and can optimize desired properties on various structures [32, 33]. One major plan is to explore flexible MOFs that have structural phase transitions using our evolutionary algorithm to optimize the heat of adsorption of them and to calculate the deliverable capacity.

Haranczyk's group developed a computational routine to build Fe(bdp) and Co(bdp) analog MOFs, in the expanded phase, working smoothly with the evolutionary algorithm. Preliminary results with this routine produced over 300 Fe(bdp) analogs successfully. However, a challenge remains to simulate the phase transitions of the structure from an expanded phase (large pore phase, or LP, in other contexts) to a collapsed phase (small pore phase, or NP, in other contexts), and vice versa. While we theoretically developed an ensemble and a corresponding equation to calculate heat of adsorption, there is a lack of appropriate force field and a lack of a computational scheme of simulation when we switch to flexible MOFs. In previous works, we

modeled methane molecules using TraPPE [112] and applied UFF [77] to simulate the interaction between methane molecules and the frameworks. Gas molecules, methane or carbon dioxide, may still be modeled as rigid or united molecules. Nevertheless, an important assumption in those works is that all frameworks are rigid, which prevents a direct application to flexible frameworks.

7.2 Force Field

Some may argue that density functional theory (DFT) could substitute for force field methods. However, DFT is generally more computationally expensive than classical force field methods, especially when we consider many possible flexible MOFs as a group. A literature search shows there are no general force fields that may be applied directly to flexible MOFs since the chemical environment is extraordinarily complex for various configurations of organic linkers and metal nodes. The general force fields such as UFF and Dreiding [113] that claim to consider nearly all atoms are too generic to accommodate complicated MOF configurations, especially ones with phase transitions, because given atom types are specifically tuned for a limited chemical environment. A force field that works for a certain MOF phase can show no validity for another phase of the same MOF. Thomas *et al.* extended the UFF by adding parameters of frequently encountered transition metal elements in MOFs [114]. Nevertheless, this extension still is designed for rigid frameworks. I have performed some molecular dynamics using UFF4MOF [114] on Fe(bdp) and MOF-5 with the general utility lattice program (GULP) [115], but eventually both structures collapsed to unrealistically condensed forms.

Force fields for particular types of MOFs have been developed using first principles to facilitate accurate molecular dynamics [116, 117]. *Ab initio* force field softwares

to generate customized force fields for given MOFs exist [118, 119, 120]. However, manual interventions are required in this type of development. These softwares have specific, limited mathematical expressions of bonded and nobonded interactions that limit their embedding in molecular system simulation packages. Nonetheless, I have tested, for example, Quick-FF on Fe(bdp) to derive a force field. The software needs a checkpoint file produced by Gaussian09 [121] frequency job, which is essentially a DFT calculation, and charges from Helpful Open-source Research TOol for N-fermion systems (HORTON) [122] atom-in-molecule analysis as input files. It did not succeed in producing convergent force field parameters. A summary of the current status of force field for flexible frameworks is that existing ones are not sufficiently robust to apply to particular MOF structures yet.

7.3 Osmotic Ensemble and Hybrid MC/MD Scheme

Besides the force field problem, to directly simulate a guest-induced structural phase transition of a flexible MOF, I need a proper thermodynamic ensemble. A literature search shows that the osmotic ensemble is the appropriate thermodynamic ensemble to apply for flexible frameworks [123, 124, 125, 126]. Recall that in Section 6.3 I proposed the $(\mu(P), N_{MOF}, P, T)$ ensemble to derive the equation for calculating heat of adsorption. This ensemble indeed is simply the osmotic ensemble by changing the stress tensor σ into external gas pressure P.

Since the osmotic ensemble is not among the commonly used ensembles in popular molecular simulation packages, it needs custom simulation methodologies. One of the methodologies that attracts our attention is the hybrid Monte Carlo and molecular dynamics (hybrid MC/MD) scheme [127]. It divides the gas adsorption in a flexible host into two stages. One is the adsorption into a rigid host treated by Monte Carlo

simulation. The other is the structural transformation of the host treated by molecular dynamics. These two stages are implemented together to approximate what physically happens. The actual implementations of hybrid MC/MD vary depending on the scope and convenience of the software package [128, 129, 130]. I note that RASPA [76] has an osmotic ensemble simulation option. However, RASPA only takes very specific mathematical forms of the force field. It implements hybrid MC/MD scheme, but the MD stage uses NVE ensemble which fixes the volume of the system. A test simulation using RASPA osmotic ensemble, UFF force filed, at 40 bar, results in a large decrease of unit cell volume of Fe(bdp) structure.

7.4 Preliminary Results

Assuming that I did not have a general methodology to develop customized force fields, I sought to reproduce others' simulation results for flexible frameworks. Literature that records simulation results of flexible MOFs with phase transition is rare, but two are found to show molecular dynamics simulation results of structural transformations of flexible MIL-53(Cr) upon CO_2 adsorption [131, 129]. They developed a particular force field with partial charges and bonded interactions of MIL-53(Cr) and validated the force field by loading given numbers of CO_2 molecules and running molecular dynamics to show the framework equilibrating in either the LP phase or the NP phase.

I set up hybrid MC/MD simulation of CO_2 adsorption in MIL-53(Cr) with the force field developed by Ghoufi *et al* [131, 129]. The CO_2 molecules are simulated using EPM model [132]. The simulation starts from the MIL-53(Cr)LP configuration at pressure of 15 bar and the temperature is set at 300 K. In each hybrid MC/MD cycle, I perform eight rounds of sequential GCMC simulation and MD simulation.

One GCMC simulation is performed in RASPA for 500 cycles to load CO_2 molecules in the framework. After GCMC simulation, three MD simulations are performed in DL_POLY_4 [133] using the NσT ensemble with timesteps of 0.00001 ps, 0.0001 ps and 0.001 ps. The NσT ensemble is essentially the isotension-isothermal ensemble where σ is the stress tensor [38]. This ensemble allows shape changes in addition to unit cell volume changes. The first two simulations run for 10,000 steps, while the last one typically runs for 100,000 steps. I implement the first two simulations to help the system to reach equilibrium. After MD simulations, one round finishes, and the program goes back to the next GCMC simulation. All the gas molecules in the system after MD simulations are inherited to the next Monte Carlo simulation. I decrease the external pressure by 1.0 bar or 0.5 bar every eight rounds depending on stability of the framework. Notice that this implementation scheme is different from the one performed by Ghoufi *et al.* [128] who implement MC and MD stages with certain probabilities. I explicitly perform MC and MD as two stages with a known number of runs.

Our preliminary results for the desorption isothermal curve are compared with experimental results [134] in Fig. 7.1. A phase transition occurs around 8 and 9 bar in the simulation. The experimental isotherms show hysteresis in the desorption and sorption curves. The desorption curve has a midpoint at 5 bar, and the sorption curve has a midpoint at 7 bar [134]. Compared with experimental isotherms, the transition pressure of the simulation desorption results are shifted about 2 bar higher. The discrepancy between simulation and experiment is likely due to deficiencies in the force field. Tuning the implementation scheme may provide finer simulation results.

The volume change of MIL-53(Cr) is presented in Fig. 7.2. Experiment shows that the LP phase MIL-53(Cr) has a unit cell volume of 1486.2 $Å^3$ at about 4.5 bar and

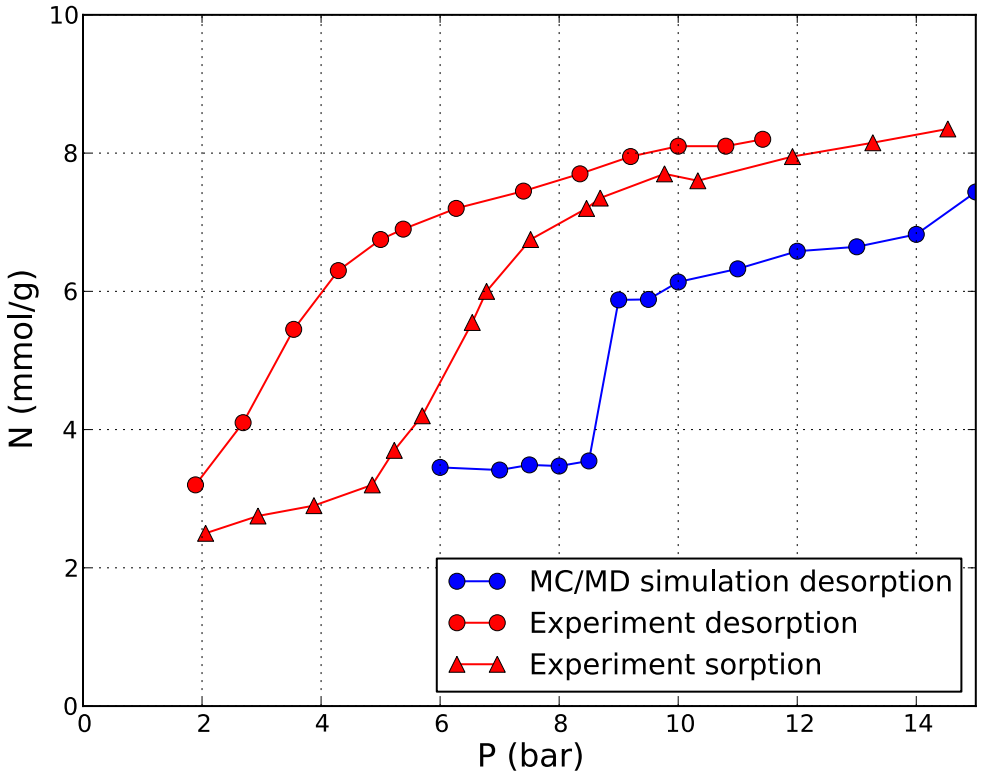


Figure 7.1 : Isotherm of MIL-53(Cr) using a hybrid MC/MD simulation scheme. The experimental isotherms at 304 K show a hysteresis, indicating a first-order phase transition. A first-order phase transition is clearly observed in the hybrid MC/MD simulation. The transition pressure is shifted from the experimental results. The simulation is performed starting from 15 bar.

the NP phase has one of 1012.8 Å^3 at about 14 bar [134]. In our simulation, the LP phase has an equilibrium volume of 1415 Å^3 and the NP phase has an equilibrium volume of 1080 Å^3, both of which agree with the experiment. This result also shows a first order structural phase transition at 9 bar.

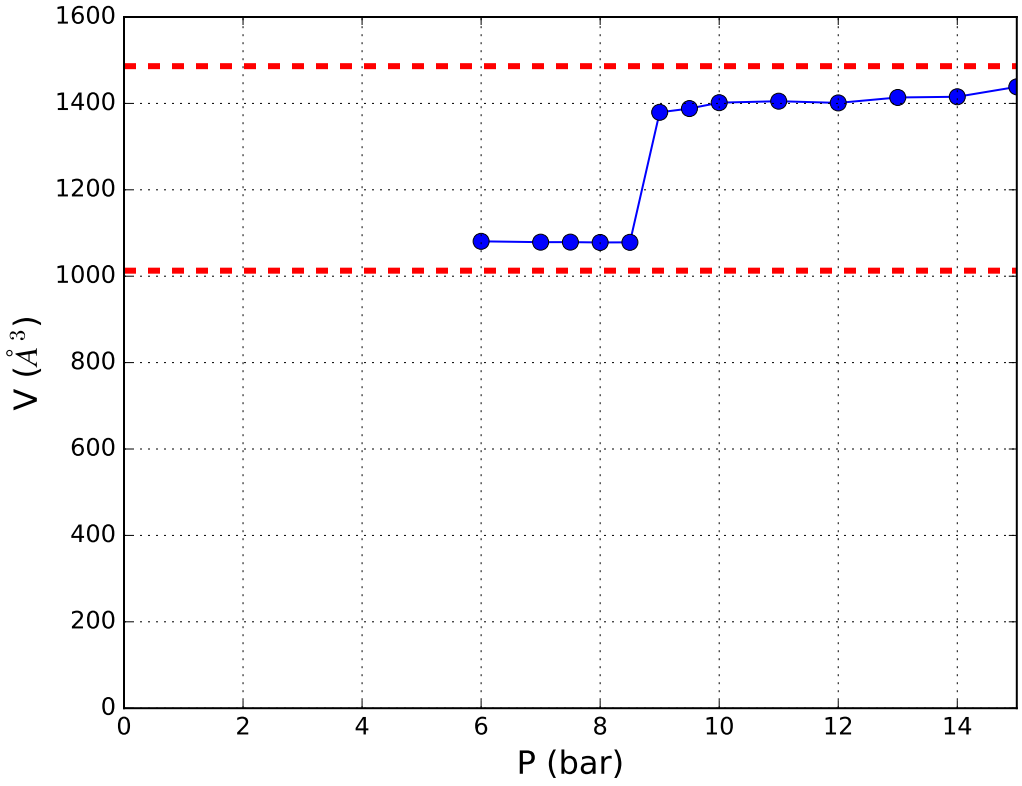


Figure 7.2 : Volume of MIL-53(Cr) as a function of pressure using the hybrid MC/MD simulation scheme. A phase transition occurs at 9 bar. The simulation is performed starting from 15 bar. Dashed lines show the experimental value of the volume in the LP and NP phases.

The current hybrid MC/MD simulation qualitatively produces the experimental result with a first-order phase transition. This methodology appears to be the applicable one after a broad survey of different combinations of ensembles, molecular

simulation softwares, and force fields. The simulations in the osmotic ensemble with DL_POLY and RASPA using a set of specific force field for MIL-53(Cr) [131, 129] is successful. With this methodology, further efforts are needed to achieve results for the phase transitions in Fe(bdp) and Co(bdp) analogs that we are ultimately interested in. And a more accurate evaluation of heat of adsorption may be performed subsequently.

7.5 Free Energy Barrier and Wang-Landau Method

The hybrid MC/MD scheme is very promising. Obstacles remain to derive a general force field for flexible MOFs. One may also seek to understand the free energy profile of the framework transition. The free energy profile determines the stable states of the system and how easy it is to transit from one state to another. MOFs with a smooth free energy landscape may experience easier phase transition than ones with high free energy barriers. This problem has been considered before [135], in which free energy as a function of unit cell volume is investigated for MIL-53(Al). Figure 6 of [135] presents two local minima of MIL-53(Al). The LP to NP phase transition of flexible MOFs should in general show a similar free energy profile with two stable states.

The number of gas molecules adsorbed can be derived from the free energy as a function of volume using the so-called Wang-Landau algorithm [136, 137]. This algorithm changes the acceptance probabilities of the Monte Carlo moves to allow all states to be visited. The Wang-Landau algorithm can be used to calculate the free

energy landscape $\Omega_{WL}(N;\mu,V)$ in the (V, N) space, which is given by

$$\begin{aligned}\Omega_{WL}(N,V;\mu,\sigma) &= F_{host}(V) + \sigma V + \Omega_{WL}(N;\mu,V) \\ &= G_{host}(V) + \Omega_{WL}(N;\mu,V)\end{aligned} \tag{7.1}$$

where σ is the stress tensor, $F_{host}(V)$ is Helmholtz free energy of the host, and $G_{host}(V)$ is the Gibbs free energy of the host that may be obtained from sampling the empty framework at a given pressure P [123]. Figure 6 of the article [137] illustrates how loading may be obtained from a free energy landscape.

7.6 Umbrella Sampling and Weighted Histogram Analysis Method (WHAM)

I consider to use the Wang-Landau algorithm to sample MOF unit cell fluctuations to obtain the free energy profile. I note that Rogge *et al.* presented an uncommon $(N,P,\sigma_a=0,T)$ ensemble to sample configurations of frameworks at different volumes [135]. Instead, I attempted a free NPT molecular dynamics to sample configurations of MIL-53(Cr) starting from the LP phase. As expected, the framework quickly reached the equilibrium at the NP phase. Sampling was basically trapped in the NP phase with small unit cell volumes. This result suggests that MIL-53(Cr) has a high free energy barrier between the NP phase and the LP phase. Perhaps this result also indicates that the NP phase has a lower free energy.

One method to circumvent this problem of trapping is to perform umbrella sampling [138]. This method samples configurations by adding a biased potential to guarantee ergodicity. It is usually performed by dividing the range of collective variable into windows and sampling each window with a particular biased potential. The

results of different windows are then combined using the weighted histogram analysis method (WHAM) [139, 140]. Despite no results shown in this thesis, I would like to point out that DL_POLY supports a native plug-in, PLUMED, which implements umbrella sampling on the fly [141]. This may be a good starting point if further work will be performed following this logic.

Chapter 8

Conclusion and Outlook

This thesis summarizes my project of computationally discovering metal-organic frameworks with high gas deliverable capacity in the Ph.D. program of physics and astronomy department of Rice University.

We have developed an evolutionary algorithm to explore metal-organic frameworks by optimizing gas deliverable capacity or accessible internal surface area [32, 33]. To seek for MOFs with high methane deliverable capacity as targeted by Department of Energy, we apply the algorithm and find several hundred new MOFs that exhibit superior methane deliverable capacity compared to the widely discussed MOF-5 and HKUST-1 in several different networks. By optimizing MOFs with high methane deliverable capacity at the 65–5.8 bar condition in nine networks, we find 48 MOFs that surpass the performance of MOF-5. The best candidate has a deliverable capacity of 191.1 v(STP)/v [32].

In a more comprehensive work, we apply the algorithm to optimize surface area of new MOFs and investigate the correlation between deliverable capacity at the 65–5.8 bar condition, deliverable capacity at the 35–5.8 bar condition, and surface area [33]. We obtain three sets of results. In the results of optimizing deliverable capacity at the 65–5.8 bar condition, the best candidate had 197.9 v(STP)/v, 11.4% higher than MOF-5. In the results of optimizing deliverable capacity at the 35–5.8 bar condition, we find 321 MOFs that have higher deliverable capacity than HKUST-1. The best candidate had 148.1 v(STP)/v, 16.8% higher than HKUST-1. In the

results of optimizing surface area, we conclude that the correlation between deliverable capacities and surface area is network dependent. While deliverable capacity defined at one metric is generally positively correlated with deliverable capacity defined at another metric, deliverable capacity and surface area are less well correlated. This methodology may be applied to efficiently explore MOFs in other networks, MOFs with multiple types of SUBs or linkers per structure, and other groups of chemical compounds.

All the results echo a recent survey of a spectrum of various materials in terms of deliverable capacity at the 65–5.8 bar condition, where none has surpassed the 200 v(STP)/v limit [2]. The present results provide another evidence that with rigid frameworks, it is difficult to reach the goal of 315 v(STP)/v deliverable capacity at the 65–5.8 bar condition set by DOE.

We are interested in recently synthesized flexible MOFs, i.e. Fe(bdp) and Co(bdp) analogs [46, 47]. These MOFs show an intriguing phase transition when the external gas pressure changes. They exhibit a heat management property that is often overlooked in the development of gas storage materials. Also, these materials show the potential of having superior deliverable capacity than typical rigid MOFs. In fact, flexibility in nanoporous materials has been observed for decades and is attracting more attention in the field of computational chemistry because of potentially finer characteristics and an even greater tunability of flexible materials [45]. I found that the osmotic ensemble, hybrid MC/MD scheme, and DL_POLY and RASPA softwares can simulate gas adsorption in flexible MOFs. I derived equations for calculating heat of adsorption in flexible frameworks. Preliminary results demonstrate that these methods are practical, and phase transition can be simulated. A variety of interesting challenges remain. One of the biggest challenges is to develop customized force

fields or even generic force fields that are suitable for simulating flexible MOFs. The force field determines the accuracy and reliability of energy minimization, molecular dynamics and Monte Carlo, and, as a consequence, needs special consideration. Another issue is how to efficiently implement the osmotic ensemble. The hybrid MC/MD scheme appears useful, but it is definitely possible to improve the implementation. One may also consider other simulation protocols that reflect what physically happens in gas adsorption. A further issue for the future is how to scale the discovery pipeline to accommodate a larger number of flexible MOF candidates. This requires the force field to deal with universal chemical environment in which MOFs reside, and also other stages of simulation to complete in a reasonable time limit using the evolutionary algorithm.

Experimental progress in the synthesis of novel types of MOFs keeps broadening our understanding of nanoporous crystalline materials. One of them is the discovery of so-called multivariate metal-organic frameworks (MTV-MOFs) in recent years by Yaghi's group, which introduces the possibility of multiple types of linker molecules and multiple functionalities in one framework [85, 142]. Multiple types of linkers in one structure may create synergy and lead to superior performance compared to MOFs with a single type of linker. Computational efforts on MTV-MOFs will likely uncover important thermodynamics and formation mechanisms of organic and inorganic components. As a conclusion, experimental synthesis and computational simulations, the two pillars of making scientific progress, will continue supporting the house of the nanoporous materials genome.

Bibliography

[1] O. M. Yaghi, G. Li, and H. Li, "Selective binding and removal of guests in a microporous metal-organic framework," *Nature*, vol. 378, pp. 703–706, 1995.

[2] C. M. Simon, J. Kim, D. A. G.-G. on, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr, and B. Smit, "The materials genome in action: Identifying the performance limits for methane storage," *Energy & Envi. Sci.*, 2015. DOI: 10.1039/c4ee03515a.

[3] P. Falcaro, R. Ricco, C. M. Doherty, K. Liang, A. J. Hill, and M. J. Styles, "MOF positioning technology and device fabrication," *Chemical Society Reviews*, vol. 43, pp. 5513–5560, 2014.

[4] M. P. Suh, H. J. Park, T. K. Prasad, and D.-W. Lim, "Hydrogen storage in metal-organic frameworks," *Chem. Rev.*, vol. 112, pp. 782–835, 2012.

[5] R. B. Getman, Y.-S. Bae, C. E. Wilmer, and R. Q. Snur, "Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal-organic frameworks," *Chem. Rev.*, vol. 112, pp. 703–723, 2012.

[6] J.-R. Li, J. Sculley, and H.-C. Zhou, "Metal-organic frameworks for separations," *Chem. Rev.*, vol. 112, pp. 869–932, 2012.

[7] Z. R. Herm, B. M. Wiers, J. A. Mason, J. M. van Baten, M. R. Hudson, P. Zajdel, C. M. Brown, N. Masciocchi, R. Krishna, and J. R. Long, "Separation of

hexane isomers in a metal-organic framework with triangular channels," *Science*, vol. 340, pp. 960–964, 2013.

[8] K. A. Cychosz, R. Ahmadab, and A. J. Matzger, "Liquid phase separations by crystalline microporous polymers," *Chem. Sci.*, vol. 1, pp. 293–302, 2010.

[9] J.-Y. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S.-B. T. Nguyen, and J. T. Hupp, "Metal-organic framework materials as catalysts," *Chem. Soc. Rev.*, vol. 38, pp. 1450–1459, 2009.

[10] A. Alshammari, Z. Jiang, and K. E. Cordova, "Metal organic frameworks as emerging photocatalysts," in *Semiconductor Photocatalysis-Materials, Mechanisms and Applications*, InTech, 2016.

[11] P. Horcajada, R. Gref, T. Baati, P. K. Allan, G. Maurin, P. Couvreur, G. Férey, R. E. Morris, and C. Serre, "Metal-organic frameworks in biomedicine," *Chem. Rev.*, vol. 112, pp. 1232–1268, 2012.

[12] L. Sarkisov, "Toward rational design of metal-organic frameworks for sensing applications: Efficient calculation of adsorption characteristics in zero loading regime," *J. Chem. Phys. C*, vol. 116, pp. 3025–3033, 2012.

[13] L. E. Kreno, K. Leong, O. K. Farha, M. Allendorf, R. P. V. Duyne, and J. T. Hupp, "Metal-organic framework materials as chemical sensors," *Chem. Rev.*, vol. 112, pp. 1105–1125, 2012.

[14] P. Kumar, A. Deep, and K.-H. Kim, "Metal organic frameworks for sensing applications," *TrAC Trends in Analytical Chemistry*, vol. 73, pp. 39–53, 2015.

[15] S.-L. Li and Q. Xu, "Metal-organic frameworks as platforms for clean energy," *Energy & Environmental Science*, vol. 6, pp. 1656–1683, 2013.

[16] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi, "The chemistry and applications of metal-organic frameworks," *Science*, vol. 341(6149), p. 1230444, 2013.

[17] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, "Large-scale screening of hypothetical metal-organic frameworks," *Nature Chemistry*, vol. 4, pp. 83–89, 2012.

[18] N. Stock and S. Biswas, "Synthesis of metal-organic frameworks (MOFs): Routes to various MOF topologies, morphologies, and composites," *Chemical Reviews*, vol. 112(2), pp. 933–969, 2012.

[19] C. Dey, T. Kundu, B. P. Biswal, A. Mallick, and R. Banerjee, "Crystalline metal-organic frameworks (MOFs): Synthesis, structure and function," *Acta Crystallographica Section B*, vol. B70, pp. 3–10, 2014.

[20] H. F. Clausen, R. D. Poulsen, A. D. Bond, M.-A. S. Chevallier, and B. B. Iversen, "Solvothermal synthesis of new metal organic framework structures in the zinc-terephthalic acid-dimethyl formamide system," *Journal of Solid State Chemistry*, vol. 178, pp. 3342–3351, 2005.

[21] A. M. Joaristi, J. Juan-Alcañiz, P. Serra-Crespo, F. Kapteijn, and J. Gascon, "Electrochemical synthesis of some archetypical Zn^{2+}, Cu^{2+}, and Al^{3+} metal organic frameworks," *Crystal Growth & Design*, vol. 12(7), pp. 3489–3498, 2012.

[22] P. J. Beldon, L. Fábián, R. S. Stein, A. Thirumurugan, A. K. Cheetham, and T. Friščić, "Rapid room-temperature synthesis of zeolitic imidazolate frame-

works by using mechanochemistry," *Angewandte Chemie International Edition*, vol. 49(50), pp. 9640–9643, 2010.

[23] A. Lagashetty, V. Havanoor, S. Basavaraja, S. Balaji, and A. Venkataraman, "Microwave-assisted route for synthesis of nanosized metal oxides," *Science and Technology of Advanced Materials*, vol. 8(6), pp. 484–493, 2007.

[24] S. Dharmarathna, C. K. King'ondu, W. Pedrick, L. Pahalagedara, and S. L. Suib, "Direct sonochemical synthesis of manganese octahedral molecular sieve (oms-2) nanomaterials using cosolvent systems, their characterization, and catalytic applications," *Chemistry of Materials*, vol. 24(4), pp. 705–712, 2012.

[25] D.-W. Jung, D.-A. Yang, J. Kim, J. Kim, and W.-S. Ahn, "Facile synthesis of MOF-177 by a sonochemical method using 1-methyl-2-pyrrolidinone as a solvent," *Dalton Transactions*, vol. 39(11), pp. 2883–2887, 2010.

[26] S. H. Jhung and J.-S. Chang. Patents KR 0627634; JP 4610531.

[27] J.-S. Chang, Y. Hwang, S. Jhung, J.-S. Hwang, and Y.-K. Seo. Patents KR 0803945.

[28] A. U. Czaja, N. Trukhan, and U. Müller, "Industrial applications of metal-organic frameworks," *Chemical Society Reviews*, vol. 38, pp. 1284–1293, 2009.

[29] C. Mellot-Draznieks, J. M. Newsam, A. M. Gorman, C. M. Freeman, and G. Férey, "De novo prediction of inorganic structures developed through automated assembly of secondary building units (AASBU method)," *Angew. Chem. Int. Ed.*, vol. 39, pp. 2270–2275, 2000.

[30] O. K. Farha, A. Ö. Yazaydn, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen, R. Q. Snurr, and J. T. Hupp, "De novo synthesis of a metal-organic framework material featuring ultrahigh surface area and gas storage capacities," *Nature Chemistry*, vol. 2, pp. 944–948, 2010.

[31] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, and R. Q. Snurr, "Computation-ready, experimental metal-organic frameworks: A tool to enable high-throughput screening of nanoporous crystals," *Chemistry of Materials*, vol. 26(21), pp. 6185–6192, 2014.

[32] Y. Bao, R. L. Martin, C. M. Simon, M. Haranczyk, B. Smit, and M. W. Deem, "In silico discovery of high deliverable capacity metal-organic frameworks," *Journal of Physical Chemistry C*, vol. 119(1), pp. 186–195, 2015.

[33] Y. Bao, R. L. Martin, M. Haranczyk, and M. W. Deem, "In silico prediction of MOFs with high deliverable capacity or internal surface area," *Physical Chemistry Chemical Physics*, vol. 17(18), pp. 11962–11973, 2015.

[34] Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp, O. K. Farha, and R. Q. Snurr, "In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm," *Science Advances*, vol. 2(10), p. e1600909, 2016.

[35] S. P. Collins, T. D. Daff, S. S. Piotrkowski, and T. K. Woo, "Materials design by evolutionary optimization of functional groups in metal-organic frameworks," *Science Advances*, vol. 2(11), p. e1600954, 2016.

[36] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, and T. K. Woo, "Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO_2 capture," *The Journal of Physical Chemistry Letters*, vol. 5(17), pp. 3056–3060, 2014.

[37] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, "Machine-learning-assisted materials discovery using failed experiments," *Nature*, vol. 533, pp. 73–76, 2016.

[38] B. Smit and D. Frenkel, *Understanding Molecular Simulation: From Algorithms to Applicatons*. San Diego: Academic Press, 2nd ed., 2012.

[39] P. S. Kutchukian and E. I. Shakhnovich, "De novo design: Balancing novelty and confined chemical space," *Expert Opin. Drug Discov.*, vol. 5, pp. 789–812, 2010.

[40] Y. Peng, V. Krungleviciute, I. Eryazici, J. T. Hupp, O. K. Farha, and T. Yildirim, "Methane storage in metal-organic frameworks: Current records, surprise findings, and challenges," *J. Am. Chem. Soc.*, vol. 135, pp. 11887–11894, 2013.

[41] K. Konstas, T. Osl, Y. Yang, M. Batten, N. Burke, A. J. Hill, and M. R. Hill, "Methane storage in metal organic frameworks," *Journal of Materials Chemistry*, vol. 22, pp. 16698–16708, 2012.

[42] M. Bosch, M. Zhang, and H.-C. Zhou, "Increasing the stability of metal-organic frameworks," *Advances in Chemistry*, vol. 2014, 2014.

[43] S. Bourrelly, P. L. Llewellyn, C. Serre, F. Millange, T. Loiseau, and G. Férey, "Different adsorption behaviors of methane and carbon dioxide in the isotypic nanoporous metal terephthalates MIL-53 and MIL-47," *Journal of the American Chemical Society*, vol. 127(39), pp. 13519–13521, 2005.

[44] L. Chen, J. P. S. Mowat, D. Fairen-Jimenez, C. A. Morrison, S. P. Thompson, P. A. Wright, and T. Düren, "Elucidating the breathing of the metal-organic framework MIL-53(Sc) with ab initio molecular dynamics simulations and in situ X-ray powder diffraction experiments," *Journal of the American Chemical Society*, vol. 135(42), pp. 15763–15773, 2013.

[45] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, and R. A. Fischer, "Flexible metal-organic frameworks," *Chemical Society Reviews*, vol. 43, pp. 6062–6096, 2014.

[46] J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, P. L. Llewellyn, N. Masciocchi, and J. R. Long, "Methane storage in flexible metal-organic frameworks with intrinsic thermal management," *Nature*, vol. 527, pp. 357–361, 2015.

[47] M. K. Taylor, T. Runčevski, J. Oktawiec, M. I. Gonzalez, R. L. Siegelman, J. A. Mason, J. Ye, C. M. Brown, and J. R. Long, "Tuning the adsorption-induced phase change in the flexible metal-organic framework Co(bdp)," *Journal of the American Chemical Society*, vol. 138(45), pp. 15019–15026, 2016.

[48] M. Khalfaoui, S. Knani, M. Hachicha, and A. Lamine, "New theoretical expressions for the five adsorption type isotherms classified by BET based on statis-

tical physics treatment," *Journal of Colloid and Interface Science*, vol. 263(2), pp. 350–356, 2003.

[49] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi, "Design and synthesis of an exceptionally stable and highly porous metal-organic framework," *Nature*, vol. 402, pp. 276–279, 1999.

[50] M. Eddaoudi, D. B. Moler, H. Li, B. Chen, T. M. Reineke, M. O'Keeffe, and O. M. Yaghi, "Modular chemistry: Secondary building units as a basis for the design of highly porous and robust metal-organic carboxylate frameworks," *Accounts of Chemical Research*, vol. 34(4), pp. 319–330, 2001.

[51] J. Kim, B. Chen, T. M. Reineke, H. Li, M. Eddaoudi, D. B. Moler, M. O'Keeffe, and O. M. Yaghi, "Assembly of metal-organic frameworks from large organic and inorganic secondary building units: New examples and simplifying principles for complex structures," *Journal of the American Chemical Society*, vol. 123(34), pp. 8239–8247, 2001.

[52] A. Schoedel, Z. Ji, and O. M. Yaghi, "The role of metal-organic frameworks in a carbon-neutral energy cycle," *Nature Energy*, vol. 1, p. 16034, 2016.

[53] B. Dudley, "BP statistical review of world energy 2015," 2015.

[54] Y. He, W. Zhou, G. Qian, and B. Chen, "Methane storage in metal-organic frameworks," *Chemical Society Reviews*, vol. 43, pp. 5657–5678, 2014.

[55] J.-R. Li, Y. Ma, M. C. McCarthy, J. Sculley, J. Yu, H.-K. Jeong, P. B. Balbuena, and H.-C. Zhou, "Carbon dioxide capture-related gas adsorption and separation in metal-organic frameworks," *Coordination Chemistry Reviews*, vol. 255, pp. 1791–1823, 2011.

[56] K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T.-H. Bae, and J. R. Long, "Carbon dioxide capture in metal-organic frameworks," *Chem. Rev.*, vol. 112, pp. 724–781, 2012.

[57] R. Sabouni, H. Kazemian, and S. Rohani, "Carbon dioxide capturing technologies: A review focusing on metal organic framework materials (MOFs)," *Environmental Science and Pollution Research*, vol. 21(8), pp. 5427–5449, 2014.

[58] E. L. First and C. A. Floudas, "MOFomics: Computational pore characterization of metal-organic frameworks," *Microporous and Mesoporous Materials*, vol. 165, pp. 32–39, 2013.

[59] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson, "The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information," *Acta Crystallographica Section B*, vol. B35, pp. 2331–2339, 1979.

[60] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quiros, N. R. Serebryanaya, P. Moeck, R. T. Downs, and A. Le Bail, "Crystallography Open Database (COD): An open-access collection of crystal structures and platform for world-wide collaboration," *Nucleic acids research*, vol. 40, pp. D420–D427, 2012.

[61] O. Delgado-Friedrichs and M. OKeeffe, "Crystal nets as graphs: Terminology and definitions," *Journal of Solid State Chemistry*, vol. 178(8), pp. 2480–2485, 2005.

[62] M. O'Keeffe, M. A. Peskov, S. J. Ramsden, and O. M. Yaghi, "The Reticular

Chemistry Structure Resource (RCSR) database of, and symbols for, crystal nets," *Accounts of Chemical Research*, vol. 41(12), pp. 1782–1789, 2008.

[63] O. Delgado-Friedrichs, M. O'Keeffe, and O. M. Yaghi, "Taxonomy of periodic nets and the design of materials," *Physical Chemistry Chemical Physics*, vol. 9, pp. 1035–1043, 2007.

[64] A. J. Howarth, Y. Liu, P. Li, Z. Li, T. C. Wang, J. T. Hupp, and O. K. Farha, "Chemical, thermal and mechanical stabilities of metal-organic frameworks," *Nature Reviews Materials*, vol. 1, p. 15018, 2016.

[65] S. J. Ramsden, V. Robins, and S. T. Hydea, "Three-dimensional Euclidean nets from two-dimensional hyperbolic tilings: Kaleidoscopic examples," *Acta Crystallographica Section A*, vol. 65, pp. 81–108, 2009.

[66] S. Hyde, O. D. Friedrichs, S. Ramsden, and V. Robins, "Towards enumeration of crystalline frameworks: The 2D hyperbolic approach," *Solid State Sciences*, vol. 8(7), pp. 740–752, 2006.

[67] G. Jones, "Genetic and evolutionary algorithms," *Encyclopedia of Computational Chemistry*, vol. 2, pp. 1127–1136, 1998.

[68] F. Daeyaert, M. De Jonge, L. Koymans, and M. Vinkers, "An ant algorithm for the conformational analysis of flexible molecules," *Journal of Computational Chemistry*, vol. 28, pp. 890–898, 2007.

[69] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk, "Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials," *Microporours and Mesoporous Materials*, vol. 149, pp. 134–141, 2012.

[70] R. L. Martin and M. Haranczyk, "Construction and characterization of structure models of crystalline porous polymers," *Cryst. Growth Des.*, vol. 14(5), pp. 2431–2440, 2014.

[71] D. R. Flower, "On the properties of bit string-based measures of chemical similarity," *J. Chem. Inf. Comput. Sci.*, vol. 38, pp. 379–386, 1998.

[72] R. Pophale, F. Daeyaert, and M. W. Deem, "Computational prediction of chemically synthesizable organic structure directing agents for zeolites," *J. Mater. Chem. A*, vol. 1, pp. 6750–6760, 2013.

[73] H. M. Vinkers, M. R. de Jonge, F. F. D. Daeyaert, J. Heeres, L. M. H. Koymans, J. H. van Lenthe, P. J. Lewi, H. Timmerman, K. V. Aken, and P. A. J. Janssen, "SYNOPSIS: SYNthesize and OPtimize System in Silico," *J. Med. Chem.*, vol. 46(13), pp. 2765–2733, 2003.

[74] J. Liu, B. Lukose, O. Shekhah, H. K. Arslan, P. Weidler, H. Gliemann, S. Bräse, S. Grosjean, A. Godt, X. Feng, K. Müllen, I.-B. Magdau, T. Heine, and C. Wöll, "A novel series of isoreticular metal organic frameworks: Realizing metastable structures by liquid phase epitaxy," *Sci. Rep.*, vol. 2, p. 921, 2012.

[75] J. W. Ponder and F. M. Richards, "An efficient Newton-like method for molecular mechanics energy minimization of large molecules," *Journal of Computational Chemistry*, vol. 8, pp. 1016–1024, 1987.

[76] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr, "RASPA: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials," *Molecular Simulation*, vol. 42(2), pp. 81–101, 2016.

[77] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, III, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *Journal of the American Chemical Society*, vol. 114, pp. 10024–10035, 1992.

[78] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, and O. M. Yaghi, "Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage," *Science*, vol. 295, pp. 469–472, 2002.

[79] J. A. Mason, M. Veenstra, and J. R. Long, "Evaluating metal-organic frameworks for natural gas storage," *Chemical Science*, vol. 5, pp. 32–51, 2014.

[80] C. M. Simon, J. Kim, L.-C. Lin, R. L. Martin, M. Haranczyk, and B. Smit, "Optimizing nanoporous materials for gas storage," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 5499–5513, 2014.

[81] R. Pophale, F. Daeyaert, and M. W. Deem, "A database of new zeolite-like materials," *Phys. Chem. Chem. Phys*, vol. 13, pp. 12407–12412, 2011.

[82] B. Li, H.-M. Wen, H. Wang, H. Wu, M. Tyagi, T. Yildirim, W. Zhou, and B. Chen, "A porous metal-organic framework with dynamic pyrimidine groups exhibiting record high methane storage working capacity," *J. Am. Chem. So.*, vol. 136, pp. 6207–6210, 2014.

[83] U. Mueller, M. Schubert, F. Teich, H. Puetter, K. Schierle-Arndt, and J. Pastré, "Metal-organic frameworks–prospective industrial applications," *Journal of Materials Chemistry*, vol. 16, pp. 626–636, 2006.

[84] S. Kandambeth, D. B. Shinde, M. K. Panda, B. Lukose, T. Heine, and R. Banerjee, "Enhancement of chemical stability and crystallinity in porphyrin-

containing covalent organic frameworks by intramolecular hydrogen bonds," *Angew. Chem. Int. Ed.*, vol. 52, pp. 13052–13056, 2013.

[85] H. Deng, C. J. Doonan, H. Furukawa, R. B. Ferreira, J. Towne, C. B. Knobler, B. Wang, and O. M. Yaghi, "Multiple functional groups of varying ratios in metal-organic frameworks," *Science*, vol. 327, pp. 846–850, 2010.

[86] L. Liu, K. Konstas, M. R. Hill, and S. G. Telfer, "Programmed pore architectures in modular quaternary metal-organic frameworks," *J. Am. Chem. Soc.*, vol. 135, pp. 17731–17734, 2013.

[87] J. E. Schmidt, M. W. Deem, and M. E. Davis, "Synthesis of a specified, silica molecular sieve using computationally predicted organic structure-directing agents," *Angew. Chem. Int. Ed.*, vol. 53, pp. 8372–8374, 2014.

[88] D. J. Tranchemontagne, J. L. Mendoza-Cortés, M. O'Keeffe, and O. M. Yaghi, "Secondary building units, nets and bonding in the chemistry of metal-organic frameworks," *Chemical Society Reviews*, vol. 38, p. 1257, 2009.

[89] B. Kesanli, Y. Cui, M. R. Smith, E. W. Bittner, B. C. Bockrath, and W. Lin, "Highly interpenetrated metal-organic frameworks for hydrogen storage," *Angewandte Chemie International Edition*, vol. 44, pp. 72–75, 2005.

[90] H.-L. Jiang, B. Liu, Y.-Q. Lan, K. Kuratani, T. Akita, H. Shioyama, F. Zong, and Q. Xu, "From metal-organic framework to nanoporous carbon: Toward a very high surface area and hydrogen uptake," *Journal of the American Chemistry Society*, vol. 133(31), pp. 11854–11857, 2011.

[91] S. Brunauer, P. H. Emmett, and E. Teller, "Adsorption of gases in multimolecular layers," *Journal of the American Chemical Society*, vol. 60(2), pp. 309–319,

1938.

[92] R. L. Martin and M. Haranczyk, "Insights into multi-objective design of Metal-Organic frameworks," *Crystal Growth & Design*, vol. 13, pp. 4208–4212, 2013.

[93] J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella, and D. J. Siegel, "Theoretical limits of hydrogen storage in Metal-Organic frameworks: Opportunities and trade-offs," *Chem. Mater.*, vol. 25, pp. 3373–3382, 2013.

[94] R. L. Martin and M. Haranczyk, "Exploring frontiers of high surface area metal-organic frameworks," *Chemical Science*, vol. 4, pp. 1781–1785, 2013.

[95] M. Eddaoudi, J. Kim, D. Vodak, A. Sudik, J. Wachter, M. O'Keeffe, and O. M. Yaghi, "Geometric requirements and examples of important structures in the assembly of square building blocks," *PNAS*, vol. 99, pp. 4900–4904, 2002.

[96] R. L. Martin, C. M. Simon, B. Medasani, D. K. Britt, B. Smit, and M. Haranczyk, "In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species," *Journal of Physical Chemistry C*, vol. 118(41), pp. 23790–23802, 2014.

[97] R. L. Martin, C. M. Simon, B. Smit, and M. Haranczyk, "In silico design of porous polymer networks: High-throughput screening for methane storage materials," *Journal of Physical Chemistry C*, vol. 136(13), pp. 5006–5022, 2014.

[98] S. Horike, S. Shimomura, and S. Kitagawa, "Soft porous crystals," *Nature Chemistry*, vol. 1, pp. 695–704, 2009.

[99] N. Lock, Y. Wu, M. Christensen, L. J. Cameron, V. K. Peterson, A. J. Bridgeman, C. J. Kepert, and B. B. Iversen, "Elucidating negative thermal expansion

in MOF-5," *The Journal of Physical Chemistry C*, vol. 114(39), pp. 16181–16186, 2010.

[100] S. S.-Y. Chui, S. M.-F. Lo, J. P. H. Charmant, A. G. Orpen, and I. D. Williams, "A chemically functionalizable nanoporous material $[Cu_3(TMA)_2(H_2O)_3]_n$," *Science*, vol. 283(5405), pp. 1148–1150, 1999.

[101] C. Serre, F. Millange, C. Thouvenot, M. Noguès, G. Marsolier, D. Louër, and G. Férey, "Very large breathing effect in the first nanoporous chromium(III)-based solids: MIL-53 or $Cr^{III}(OH)\cdot\{O_2C\text{-}C_6H_4\text{-}CO_2\}\cdot\{HO_2C\text{-}C_6H_4\text{-}CO_2H\}_x\cdot H_2O_y$," *Journal of the American Chemical Society*, vol. 124(45), pp. 13519–13526, 2002.

[102] Y. Ye, X. Wu, Z. Yao, L. Wu, Z. Cai, L. Wang, X. Ma, Q.-H. Chen, Z. Zhang, and S. Xiang, "Metal-organic frameworks with a large breathing effect to host hydroxyl compounds for high anhydrous proton conductivity over a wide temperature range from subzero to 125 °C," *Journal of Materials Chemistry A*, vol. 4, pp. 4062–4070, 2016.

[103] C. Mellot-Draznieks, C. Serre, S. Surblé, N. Audebrand, and G. Férey, "Very large swelling in hybrid frameworks: A combined computational and powder diffraction study," *Journal of the American Chemical Society*, vol. 127(46), pp. 16273–16278, 2005.

[104] K. W. Chapman, G. J. Halder, and P. J. Chupas, "Pressure-induced amorphization and porosity modification in a metal-organic framework," *Journal of the American Chemical Society*, vol. 131(48), pp. 17546–17547, 2009.

[105] S. Sircar, "Pressure swing adsorption: research needs by industry," *Fundamentals of Adsorption*, pp. 815–843, 1991.

[106] A. L. M. Fokion. Karavias, "Isosteric heats of multicomponent adsorption: Thermodynamics and computer simulations," *Langmuir*, vol. 7(12), pp. 3118–3126, 1991.

[107] A. L. Mayers, "Thermodynamics of adsorption in porous materials," *AlChE Journal*, vol. 48(1), pp. 145–160, 2002.

[108] G. N. Lewis, M. Randall, K. S. Pitzer, and L. Brewer, *Thermodynamics*. McGraw-Hill, 1961.

[109] G. B. Woods, A. Z. Panagiotopoulos, and J. S. Rowlinson, "Adsorption of fluids in model zeolite cavities," *Molecular Physics*, vol. 63(1), pp. 49–63, 1988.

[110] A. Lyubchyk, I. A. A. C. Esteves, F. J. A. L. Cruz, and J. P. B. Mota, "Experimental and theoretical studies of supercritical methane adsorption in the MIL-53(Al) metal organic framework," *The Journal of Physical Chemistry C*, vol. 115(42), pp. 20628–20638, 2011.

[111] L. Cao and Z. Zhou, *Thermology, Thermodynamics and Statistical Physics*. Science Press, 2008.

[112] M. G. Martin and J. I. Siepmann, "Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes," *The Journal of Physical Chemistry B*, vol. 102(14), pp. 2569–2577, 1998.

[113] S. L. Mayo, B. D. Olafson, and W. A. Goddard, "DREIDING: A generic force field for molecular simulations," *The Journal of Physical Chemistry*, vol. 94(26),

pp. 8897–8909, 1990.

[114] M. A. Addicoat, N. Vankova, I. F. Akter, and T. Heine, "Extension of the universal force field to metal-organic frameworks," *Journal of Chemical Theory and Computation*, vol. 10(2), pp. 880–891, 2014.

[115] J. D. Gale and A. L. Rohl, "The general utility lattice program (GULP)," *Molecular Simulation*, vol. 29, pp. 291–341, 2003.

[116] S. Amirjalayer, M. Tafipolsky, and R. Schmid, "Exploring network topologies of copper paddle wheel based metal-organic frameworks with a first-principles derived force field," *The Journal of Physical Chemistry C*, vol. 115(31), pp. 15133–15139, 2011.

[117] L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, and V. V. Speybroeck, "Ab initio parametrized force field for the flexible metal-organic framework MIL-53(Al)," *Journal of Chemical Theory and Computation*, vol. 8(9), pp. 3217–3231, 2012.

[118] M. Tafipolsky and R. Schmid, "Systematic first principles parameterization of force fields for metalorganic frameworks using a genetic algorithm approach," *The Journal of Physical Chemistry B*, vol. 113(5), pp. 1341–1352, 2009.

[119] S. Bureekaew and R. Schmid, "Hypothetical 3D-periodic covalent organic frameworks: Exploring the possibilities by a first principles derived force field," *CrystEngComm*, vol. 15, pp. 1551–1562, 2013.

[120] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. V. Speybroeck, "QuickFF: A program for a quick and easy derivation

of force fields for metal-organic frameworks from ab initio input," *Journal of Computational Chemistry*, vol. 36, pp. 1015–1027, 2015.

[121] M. Frisch *et al.*, "Gaussian09." Gaussian Inc., Wallingford, CT, USA, 2009.

[122] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, K. Boguslawski, M. Chan, Y. Zhao, T. D. Kim, S. Vandenbrande, D. Yang, C. E. González-Espinoza, S. Fias, P. A. Limacher, D. Berrocal, A. Malek, and P. W. Ayers, "Horton," 2015.

[123] F.-X. Coudert, M. Jeffroy, A. H. Fuchs, A. Boutin, and C. Mellot-Draznieks, "Thermodynamics of guest-induced structural transitions in hybrid organic-inorganic frameworks," *Journal of the American Chemical Society*, vol. 130(43), pp. 14294–14302, 2008.

[124] M. Jeffroy, A. H. Fuchs, and A. Boutin, "Structural changes in nanoporous solids due to fluid adsorption: Thermodynamic analysis and monte carlo simulations," *Chemical Communications*, pp. 3275–3277, 2008.

[125] J. Zang, S. Nair, and D. S. Sholl, "Osmotic ensemble methods for predicting adsorption-induced structural transitions in nanoporous materials using molecular simulations," *The Journal of Chemical Physics*, vol. 134, p. 184103, 2011.

[126] F.-X. Coudert, A. Boutin, M. Jeffroy, C. Mellot-Draznieks, and A. H. Fuchs, "Thermodynamic methods and models to study flexible metal-organic frameworks," *ChemPhysChem*, vol. 12(2), pp. 247–258, 2011.

[127] S. V. Nedea, A. J. H. Frijns, A. A. van Steenhoven, A. J. Markvoort, and P. A. J. Hilbers, "Hybrid method coupling molecular dynamics and Monte Carlo simulations to study the properties of gases in microchannels and nanochannels," *Physical Review E*, vol. 72, p. 016705, 2005.

[128] A. Ghoufi and G. Maurin, "Hybrid Monte Carlo simulations combined with a phase mixture model to predict the structural transitions of a porous metal-organic framework material upon adsorption of guest molecules," *The Journal of Physical Chemistry C*, vol. 114(14), pp. 6496–6502, 2010.

[129] A. Ghoufi, A. Subercaze, Q. Ma, P. Yot, Y. Ke, I. Puente-Orench, T. Devic, V. Guillerm, C. Zhong, C. Serre, G. Frey, and G. Maurin, "Comparative guest, thermal, and mechanical breathing of the porous metal organic framework MIL-53(Cr): A computational exploration supported by experiments," *The Journal of Physical Chemistry C*, vol. 116(24), pp. 13289–13295, 2012.

[130] L. Zhang, Z. Hu, and J. Jiang, "Sorption-induced structural transition of zeolitic imidazolate framework-8: A hybrid molecular simulation study," *Journal of the American Chemical Society*, vol. 135(9), pp. 3722–3728, 2013.

[131] F. Salles, A. Ghoufi, G. Maurin, R. G. Bell, C. Mellot-Draznieks, and G. Férey, "Molecular dynamics simulations of breathing MOFs: Structural transformations of MIL-53(Cr) upon thermal activation and CO2 adsorption," *Angewandte Chemie*, vol. 120(44), pp. 8484–8491, 2008.

[132] J. G. Harris and K. H. Yung, "Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model," *The Journal of Physical Chemistry*, vol. 99(31), pp. 12021–12024, 1995.

[133] I. Todorov, W. Smith, and Cheshire, "The DL POLY 4 user manual," *STFC, STFC Daresbury Laboratory, Daresbury, Warrington, Cheshire, WA4 4AD, United Kingdom, version 4.0*, 2011.

[134] C. Serre, S. Bourrelly, A. Vimont, N. A. Ramsahye, G. Maurin, P. L. Llewellyn,

M. Daturi, Y. Filinchuk, O. Leynaud, P. Barnes, and G. Férey, "An explanation for the very large breathing effect of a metal-organic framework during CO_2 adsorption," *Advanced Materials*, vol. 19(17), pp. 2246–2251, 2007.

[135] S. Rogge, L. Vanduyfhuys, A. Ghysels, M. Waroquier, T. Verstraelen, G. Maurin, and V. V. Speybroeck, "A comparison of barostats for the mechanical characterization of metal-organic frameworks," *Journal of Chemical Theory and Computation*, vol. 11(12), pp. 5583–5597, 2015.

[136] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letters*, vol. 86(10), pp. 2050–2053, 2001.

[137] D. Bousquet, F.-X. Coudert, and A. Boutin, "Free energy landscapes for the thermodynamic understanding of adsorption-induced deformations and structural transitions in porous materials," *The Journal of Chemical Physics*, vol. 137, p. 044118, 2012.

[138] G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of Computational Physics*, vol. 23(2), pp. 187–199, 1977.

[139] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method," *Journal of Computational Chemistry*, vol. 13(8), pp. 1011–1021, 1992.

[140] M. Souaille and B. Roux, "Extension to the weighted histogram analysis

method: Combining umbrella sampling with free energy calculations," *Computer Physics Communications*, vol. 135, pp. 40–57, 2001.

[141] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," *Computer Physics Communications*, vol. 180(10), pp. 1961–1972, 2009.

[142] X. Kong, H. Deng, F. Yan, J. Kim, J. A. Swisher, B. Smit, O. M. Yaghi, and J. A. Reimer, "Mapping of functional groups in metal-organic frameworks," *Science*, vol. 341(6148), pp. 882–885, 2013.